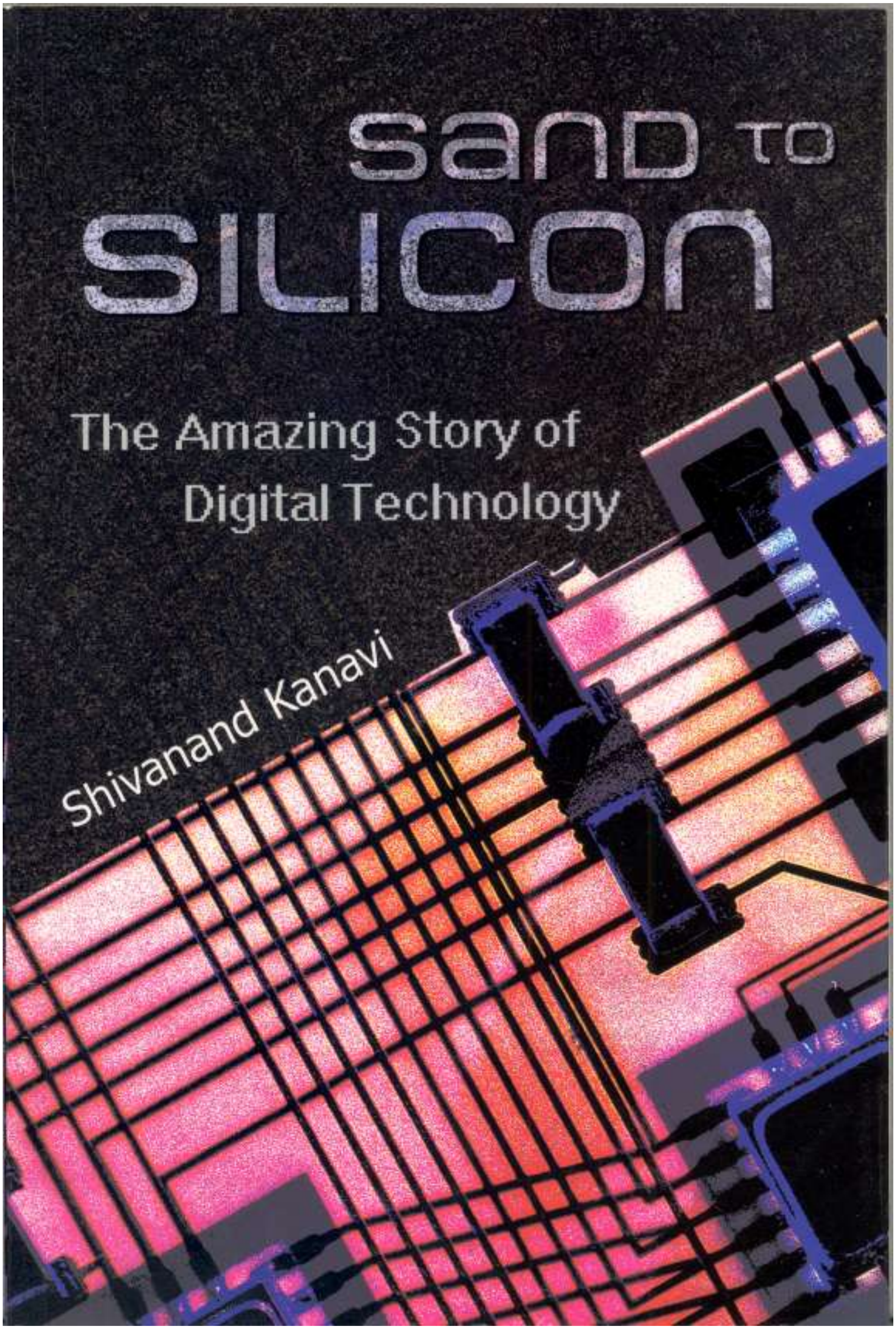


SAND TO SILICON

The Amazing Story of Digital Technology

Shivanand Kanavi

SAND TO SILICON

The amazing story of digital technology

SHIVANAND KANAVI

Photographs: Palashranjan Bhaumick

First Published by

Tata McGraw Hill 2004

Published 2006 by

Rupa & Co.

CONTENTS

Press Reviews

Mr Shivanand Kanavi's maiden book covers the entire gamut of developments in semiconductors, computers, fibre optics, telecommunications, optical technologies and the Internet, while holding a light up to the genius, individual and collective, that brought the digital dream to throbbing life.

-Deccan Herald

Repaints digital history from the perspective of the contribution of myriad brilliant Indian scientists, researchers, academicians and entrepreneurs, all of whom played a critical role in technological breakthroughs that have made IT what it is today.

-Express Computers

For someone, who would like to know how the World Wide Web came into being, or what a chip really does, Sand to Silicon has all the answers. Surprisingly easy to understand, considering the complexity of the subject. Mr Kanavi simplifies technology for the common man, using ordinary, if unusual metaphors. His book is international enough to be about technology in general, but he takes care to underscore the Indian contribution to global advances in technology.

-Financial Express

Chronicles possibly for the first time-the story from a 'desi' perspective and weaves Indian achievers and achievements into the very fabric of IT and its brief international history. Reading it, will make every Indian proud.

-The Hindu

Response

Dear Shri Shivanand Kanavi,

Thank you for sending me a copy of your book "Sand to Silicon: The amazing story of digital technology. I have gone through the book and particularly I liked the chapters "Optical Technology: Lighting up our lives" (page 178) and "Epilogue: The Collective Genius" (page 243). My best wishes.

Yours sincerely,

A.P.J.Abdul Kalam

August 9, 2004

President of India, Rashtrapati Bhavan, New Delhi. 110004

"Kanavi is a gifted writer in the mold of Isaac Assimov. He explains science and technology in a simple manner. This enables hi readers with little exposure to science to understand technology, its phenomena and processes. His book *Sand to Silicon* starts with the invention of the transistor, which led to digital electronics, integrated circuits, computers and communications. He narrates developments such that readers feel they are participating in the whole process. He also gives a human face to technology by talking about the persons behind it. Everyone who reads *Sand to Silicon,* irrespective of their background in science of arts, will get deep insight in the world of digital electronic, which had touched our lives from High Definition TVs to mobile phones."

-F C KOHLI, IT pioneer

"We are witness to the way information and Communications Technology (ICT) is revolutionizing everything around us today. Shivanand Kanavi provides a compelling and breathtaking account of the science and technology that went into this revolution, with simplicity and elegance that is the hallmark of his writings. Equally fascinating is his account of the role of the 'Indian genius' in powering the ICT revolution, with the rarest of rare insights acquired through painstaking research, this masterpiece is 'must' for everyone."

-R A MASHELKAR, FRS, Former DIRECTOR GENERAL, CSIR

"*Sand to Silicon* is elegant in its simplicity. Any non-engineer whose world is touched by micro-electronics, software and telecommunications needs to read it, because it brings understanding of these technologies within all of our reach."

-PROF CLAYTON CHRISTENSEN, HARVARD BUSINESS SCHOOL

“Presents extremely complex scientific concepts in an easy to understand manner…this book provides a good foundation of the key building blocks. Should be a required reading for all IT practitioners.”

-YASHIRO MASAMOTO, CHAIRMAN, SHINSEI BANK, JAPAN

“There is proverb in Marathi, which roughly translates into, ‘with committed efforts one can even squeeze oil out of sand’. ***Sand to Silicon*** is a saga of human ingenuity and efforts in realizing ever better results, which have made a paradigm shift in the history of human development. I would like to compliment Shivanand Kanavi for bringing out this book, which I am sure, would benefit all those readers who are interested in today’s technology revolution.”

-**ANIL KAKODKAR, Former CHAIRMAN, ATOMIC ENERGY COMMISSION**

ACKNOWLEDGEMENTS

Words in this section appear customary, but are entirely true. A project this ambitious would not have been possible without the enthusiastic help of literally, hundreds of people. However, inadequacies in the book are solely mine.

Tata Sons supported the author financially during the research and writing of the book, without which this book would not have been possible. However, the views expressed in this book are entirely those of the author and do not represent those of the Tata Group.

I acknowledge with gratefulness the contributions made by:

- R. Gopalakrishnan, of the Tata Group for championing the project through thick and thin, without whose encouragement Sand to Silicon would have remained a gleam in the author's eyes.

- S. Ramadorai of TCS for his constant encouragement and support in my efforts at communication of science and technology to lay persons and for writing a valuable Introduction to this edition ..

- F.C. Kohli for providing insights and perspective on many issues in global technology and business history.

- Ashok Advani, of Business India for mentoring me and turning a theoretical physicist and an essayist like me into a business journalist.

- Kesav Nori (TCS), Juzer Vasi (lIT,B), Ashok Jhunjhunwala (lIT,M), Bishnu Pradhan, Sorab Ghandhi, Jai Singh, Umesh Vazirani (UC Berkeley), Kannan Kasturi, and YR. Mehta (Tata Motors) for taking time off to give detailed feedback on various chapters.

- My publisher, Kapish Mehra of Rupa & Co for publishing the new edition.

- Sanjana Roy Choudhury for excellent editing of this edition.

- Satyabrata Sahu for diligently checking the proofs.

- Hundreds of experts who patiently shared their valuable time, their knowledge-base and friendship:

AV Balakrishnan (UCLA), Abhay Bhushan, Amar Bose (Bose Corp), Arogyasami Paul Raj (Stanford U), Arun Netravali, Aravind Joshi, Avtar Saini, Bala Manian (Saraswati Partners), Balaji Prabhakar (Stanford U), Basavraj Pawate (TI), Bhaskar Ramamurthy (IIT, M), Birendra Prasada, Bishnu Atal, Bishnu Pradhan, Bob Taylor,

Bobby Mitra (TI), Chandra Kudsia, C.K.N. Mangla, C.K.N. Patel (Pranalytica), C Mohan (IBM, Almaden), D.B. Phatak (IIT,B), Debasis Mitra (Bell Labs), Desh Deshpande (Sycamore

Networks), Dinesh (IIT, B), F.C Kohli (Tata Group), H. Kesavan (U of Waterloo), Jack Kilby, Jai Menon (IBM, Almaden), Jai Singh, Jayant Baliga (NC State U), Jitendra Mallik (UC Berkeley), Jnaan Dash, (Sonata Software), Juzer Vasi (IIT, B), K. Bala (TI), K. Kasturirangan (NIAS), K Mani Chandy (Caltech), Kamal Badada (TCS), Kanwal Rekhi, Kesav Nori (TCS), Keshav Parhi (U of Minnesota), Krishna Saraswat (Stanford U), Kriti Amritalingam (IIT, B), Kumar Sivarajan (Tejas Networks), Kumar Wikramasinghe (IBM, Almaden), Luv Grover (Bell Labs), M. Vidyasagar (TCS), Madhusudan (MIT), Manmohan Sondhi (Avaya),

Mathai Joseph (TRDDC), Mohan Tambay, Mriganka Sur (MIT), N. Jayant (Georgia Tech), N. Vinay (IISc), N. Yegnanarayana (IIT, M), Nambinarayanan, Nandan Nilekani (Infosys), Narendra Karmarkar, Narinder Singh Kapany, Naveen Jain, Neera Singh (Telecom Ventures), Niloy Dutta (U Conn), PP Vaidyanathan (Caltech), P Venkatarangan (UC San Diego), Pallab Bhattacharya (U Michigan), Pallab Chatterji (I2), Prabhu Goel, Pradeep Khosla (CMU), Pradeep Sindhu (Juniper Networks), Prakash Bhalerao, Pramod Kale, Praveen Chaudhari , R. Narasimhan, Raghavendra Cauligi (USC), Raj Reddy (CMU), Raj Singh (Sonoa Systems), Rajendra Singh (Telecom Ventures), Rajeev Motwani (Stanford U), Rajeev Sangal (IIIT, Hyd), Rakesh Agarwal (IBM, Almaden), Rakesh Lal (IIT,B), Ramalinga Raju (Satyam), Ramesh Agarwal (IBM, Almaden), Ravi Kannan (Yale), Roddam Narasimha (IISc), S. Keshav (U of Waterloo), S. Mittal (I2), Sam Pitroda, Sanjit Mitra (UC Santa Barbara), Sanjiv Sidhu (I2), Sorab Ghandhi, Subra Suresh (MIT), Timothy Gonsalves (IIT, M), Tom Kailath (Stanford U), U.R. Rao, Umesh Mishra (UC Santa Barbara), Umesh Vazirani (UC Berkeley), Upamanyu Madhow (UC Santa Barbara),

V Rajaraman (IISc), VVS. Sarma (IISc),Venky Narayanamurthy (Harvard U), Vijay Chandru (IISc), Vinay Chaitanya, Vinay Deshpande (Ncore), Vinod Khosla (KPCB), Vijay Madisetti (Georgia Tech), Vijay Vashee, Vinton Cerf (Google), Vivek Mehra (August Capita)), Vivek Ranadive (TIBCO), Yogen Dalal (Mayfield Ventures).

- Mike Ross at IBM and Saswato Das at Bell Labs for making interviews possible at T.J. Watson Research Centre, Yorktown Heights, Almaden Research Centre and at Bell Labs, Murray Hills.

- Christabelle Noronha, of the Tata Group for coordinating varied parts of this complex project with remarkable drive.

- T.R. Doongaji, FN. Subedar, Romit Chatterji, R.R. Shastri, K.R. Bhagat, Juthika Choksi Hariharan of the Tata Group for providing invaluable infrastructural support.

- Delphine Almeida and B. Prakash of TCS for providing library support. Elsy Dias, Fiona Pinto and Sujatha Nair for secretarial help.

- Radhakrish nan, Debabrata Paine, Raj Patil, Iqbal Singh, Anand Patil, Srinivas Rajan and innumerable friends in TCS and Tata Infotech, for their hospitality in North America.

- Raj Singh, K.V Kamath, Arjun Gupta, Arun Netravali, Kanwal Rekhi, Jai Singh, R. Mashelkar, Desh Deshpande and Jacob John for encouragement during the incubation of the project.

- Balle, Madhu, Geetha, Sanjoo, Ashok, Pradip, Revathi, Kannan and Bharat for friendship and inputs.

- Palashranjan Bhaumick for visually recording the interviews and invaluable support at various stages of the project.

- My parents, Chennaveera Kanavi and Shanthadevi Kanavi and in laws Col. Gopalan Kasturi and Lakshmi Kasturi who inspired me to become a writer.

- Last but not the least my wife Radhika and children Rahul and Usha for very generously writing off all my idiosyncrasies as due to "writing stress".

For hundreds of people in the IT Industry and academia who looked at it as their own project and gave contacts and suggestions.

Shivanand Kanavi

****

AUTHOR PROFILE

Born in Dharwad, Karnataka in a family of litterateurs, in 1953, Shivanand is a theoretical physicist from IIT Kanpur and Northeastern University, Boston and has carried out research at IIT Bombay, as well. He has researched in Super Symmetry, Quantum Gravity and Particle Physics and has published papers in international physics journals.

After a teaching and academic career, Shivanand became an economic consultant and later turned to business journalism.

Until June 2004, he was the Executive Editor of *Business India*. He has extensively covered issues in Business and Technology in his journalistic career. He was awarded the Madhu Valluri Award for IT Journalism for the year 2004.

Shivanand has authored a highly acclaimed book **"Sand to Silicon: The amazing story of digital technology"** (Rupa & Co. 2006). The book tells the story of the global development of semiconductors, microelectronics, lasers, fibre optics, optical networking, telecommunications, computing and the Internet. The book chronicled for the first time seminal contributions of Indians to these fields.

Shivanand has worked in Tata Consultancy Services as Vice President-Special Projects (2004-2010). He has edited a compilation, **“Research by Design-Innovation and TCS”** (Rupa &

Co. 2007). The book was released by President Kalam during the Silver Jubilee of TCS R&D, in February 2007.

He is currently Global Head, Marketing and Strategic Communication at CMC Ltd a TATA Enterprise and a subsidiary of TCS.

He occasionally writes in *Business India, Rediff.com* and is associated with the Quarterly magazine *Ghadar Jari Hai—The Revolt Continues* that explores Indian history, philosophy and thought material away from the jaundiced eye of Eurocentrism.

He can be reached at : +9198202 25869 (Cell)

OR

skanavi@yahoo.com; skanavi@gmail.com

shivanand@iitkalumni.org

PS: Shivanand Kanavi retired from TATA Group on Nov 1, 2013. Currently he is Adjunct Faculty at the National Institute of Advanced Studies, Bengaluru and Consulting Editor, Business India and contributes frequently to Business India, Rediff.com and Prajavani.
May 14, 2017
Blog: www.reflections-shivanand.blogspot.com/
Twitter: @shivanandkanavi
LinkedIn: https://www.linkedin.com/pub/shivanand-kanavi/27/361/330

The First Edition

FOREWORD

A hundred years ago, when India was under the colonial yoke, few people could have imagined this country playing a stellar role in the evolution of some of the world's most modern technologies. One person who did was Jamsetji Nusserwanji Tata, the visionary who laid the foundation of modern Indian industry. His work was built upon by his successors, notably

Jehangir Ratanji Dadabhoy (JRD) Tata, who reached out to new frontiers in industry and technology. Jamsetji and JRD were fired by a burning desire to exploit technology to improve the quality of life of all Indians.

Today, we live in a very different India from the one these two remarkable men were born into. The lumbering giant yoked to poverty has made way for a country standing high on the cusp of greatness. That vast reservoir of potential is finally being realised as India gets ready to claim its rightful place on the world stage.

The Indian Institute of Science, Bangalore and the Tata Institute of Fundamental Research, Mumbai are two temples of learning that have helped India make the transition from the technological backwaters to scientific significance. Jamsetji was instrumental in setting up the first of these trailblazing institutions and JRD played a pivotal role in getting the second up and running.

It happens to be 100 years since Jamsetji persuaded Charles Page Perin and the Geologist, CM Weld, to visit our country to prospect for iron ore in Central India. The very next year, in 1904, Jamsetji passed away, leaving his son Dorabji with the task of actually setting up the steel plant. JRD was born in 1904, just a few weeks after Jamsetji's death.

This book is in a sense, a commemorative tribute to these two exceptional men, the great savants who brought new technologies to India.

As we chart our path into the future, we need to cast a glance at the past—to the hopes and achievement of the many brilliant Indians who have, in the country and abroad, played a critical role in the development of the technologies powering the Information Technology and telecommunications revolution. *Sand to Silicon* is just such an attempt—an excursion into the past-to see how these technologies were developed, and what role Indian scientists and engineers played in the process.

Today, India is internationally acknowledged as a fountainhead of technology developers. It wasn’t always so. It took much hard work, fortitude and inspiration to arrive at this point of a voyage that remains far from complete. *Sand to Silicon* traces the history of the expedition that has made Information Technology and Communications central to our existence. It holds a light to the genius, individual and collective, that brought the digital dream to throbbing life.

The Tata Group's support to Shivanand Kanavi in the writing of this book is based as much on the need to chronicle, from an Indian perspective, the growth and spread of Information Technology and Communications, as it is to inform and invigorate today's generation.

We hope that this book will help to further the understanding of these modern technologies, simplifying their complexities for the reader, and to celebrate the many ways in which they operate, for improving the quality of life in India and all over the world.

The Tata Group

The Second Edition

FOREWORD

Several years ago, I was struck by the increasing gaps in understanding new technologies of computing and communications among top company leaderships. Such leaders did not need to be experts themselves, they needed to be able to deal with a band of 'experts'. They needed to be comfortable with this emerging world, granting that it would be a different world from the one in which they grew up. I wished there would be an easy-to-read, narrative styled story about the evolution of these technologies-something that did not require even a bachelor's degree in physics to understand.

About four years ago, when he was working with *Business India*, Shivanand Kanavi approached me about the idea of this book. His idea seemed to be in a similar area, though not quite. He had been a professor of physics and was a doctoral student turned journalist. He had also been lead writer of a cover story in *Business India* about the significant contributions of scientists and technologist of Indian origin. He was fascinated that such scientists had contributed to breakthrough thinking in semiconductors, electronics, chips and microelectronics, whereas the world recognised only the software skills of Indians. The idea of an easy-to-read book was not his 'natural' idea.

It took a discussion of considerably divergent views to bring his idea and mine together. He then took a year off to write this up, thanks to *Business India* and Tata Group support. The result was *Sand to Silicon.* It was well recieved I was pleased that some business associates asked for extra copies, a surrogate tribute to what they thought about the book.

I am pleased that there has been enough encouragement from the readers that *Sand to Silicon* is coming out so soon after the first edition (2004). India and Indians began to lose their economic supremacy to the west from the seventeenth century because they could no adapt well enough to the required attitudes and technologies in the times that followed. This time around, that should not happen. In a very small way, Shivanand' book may play a helpful role.

Mumbai, 19th June, 2006 **R. Gopalakrishnan**

Second Edition

INTRODUCTION

I have been offered the privilege of introducing the reader to the paperback edition of Shivanand Kanavi's book *Sand to Silicon* which takes us down the fascinating journey of the evolution of Information Technology in the twenty-first century.

While some of us have witnessed the IT revolution first hand, I truly believe that the best of IT is yet to come. Imagine a future of digital homes with devices talking to each other, programmed to our precise need all controllable on the move by a mobile phone that even ensures the film of your choice is loaded and ready to be viewed on your LCD screen when you get back from work!

If I were to crystal-gaze I see many challenges of today, becoming I realities tomorrow, I see the excited anticipation of the next big boom waiting to happen. Very simply put, more and more things will get digital (products and services). This digitisation means that a host of different devices and services will talk the same language, which then offers the exciting possibility of their convergence. The powerful combined effects of these two developments—digitisation and convergence, will continue to lead to disruptive innovations, innovations that will redefine the way we do business, the way we communicate, our choices and modes of entertainment, the way we conduct business and the way our homes are designed.

If I were to single out the defining factor in IT thus far (remember we arc still in the infancy of the IT revolution) it is undoubtedly the Internet. It has blossomed from "one to one" to "one to many" to a "many to many" form of communication making it a uniquely social medium. Today mass collaborations on the Internet, such as virtual supercomputers, weblogs and open source programs are shaking up businesses.

This exponential growth of IT as we see today, was made possible by the significant contributions of uniquely gifted people, in diverse corners of the world, each of their inventions providing a building block to the next level of growth. *Sand to Silicon's* most immediate value is in helping the reader understand these technologies, their implications, as well as the people behind them. Interwoven with anecdotes and first-person accounts, the book winds its way from early beginnings in chip and integrated circuit design to computing fundamentals and mathematical algorithms that challenge computer professionals. The story moves on to the age of personal computing, the telecom revolution, optical technologies, mobile communication and finally, the Internet. What would be of particular interest to young readers, new to the IT field, is the introduction to the men behind the scenes, many of whom are established figures today. The book reveals their early beginnings and experiments before they became luminaries of the business of IT. To youngsters India it would be highly inspiring to see many Indians as global technology pioneers. It can be safely said that this book is the first in chronicling Indian contribution to digital technology.

The book fills an obvious need for a systematic account of the account of the fascinating developments in IT, its people and the path-breaking inventions they created. Shivanand has put to good use his knowledge as theoretical physicist and his journalistic skills to demystify seemingly complex technologies.

I am sure the book will make an interesting and informative read to all those young and old, curious to familiarise themselves with the digital world—how it all began and where it's likely to take us all.

S. Ramadorai

CEO& MD, Tata Consultancy Services Ltd

PROLOGUE

Due to the success of the software industry in India, Information Technology has become synonymous with software or computers. But that is a very narrow view. Modern day IT is a product of the convergence of computing and communication technologies. It is not surprising that there is a computer within every telephone and a telephone within every computer.

The technologies that form the foundation of IT, which have made it accessible and affordable to hundreds of millions of people, are: semiconductors, microchips, lasers and fibre optics.

IT has emerged as a technology that has radically changed the old ways of doing many things-be it governance, manufacturing, banking, communicating, trading commodities and shares, or even going to the university or the public library! It has the potential to disrupt the .economic, social and political status quo.

Then why should we welcome it? Well, there are disruptions and disruptions. Disruption means drastically altering or destroying the structure of something. So whether the disruptive potential of anything is to be welcomed or opposed depends on what it disrupts: the old, the stale, the iniquitous and the oppressive; or the young, the fresh and the just.

If a technology has the potential to empower the individual, enhance his or her faculties and capabilities, then it has to be welcomed. Similarly, if a technology increases the possibilities of cooperation, collaboration or communication and break hierarchical and sectarian barriers, then, too, it should be welcomed.

However, modern Information Technology can do both. That is why the individualists like it and so do the collectivists. But the two categories have been wrongly posed in the twentieth century as opposites. Neither the individual nor any collective can claim supremacy. The individual and the collective have to harmonise relations among themselves to lead to a higher level of society. That is the message of the twenty-first century, and IT is an enabling technology for bringing about such harmony.

That is the reason I have chosen a seeming oxymoron-creative disruption-to describe the effect of IT. It will disrupt sects, cliques, power brokers and knowledge and information monopolies. It will extend the democracy that we tasted in the twentieth century to new and higher levels.

In the twenty-first century, the individual will flower, the collective will empower and IT will enable this. Mind you, I am not advocating that technology by itself will bring about a revolution. It can't; it has to be brought about by humans and no status quo can be altered without a fight.

Today, nobody can ignore IT. It is proliferating all around us. Modern cars have forty to fifty microprocessors inside them to control navigation, fuel injection, braking, suspension, entertainment, climate control and so on. Even the lowly washing machines, colour TVs and microwave ovens have chips controlling them. DVDs, VCDs, MP3 players, TV remote controls, cell phones, digital diaries, ATMs, cable T'I, the Internet, dinosaurs in movies, email, chat and so on are all products of IT.

Hence, awareness of the fascinating story of IT is becoming a necessity.

This book is a modest attempt to espouse IT's evolution, achievements, potential and intellectual challenges that have motivated some of the best minds in the world to participate in its creation.

The pervasive usefulness of IT makes us curious to go behind the boxes-PCs and modems-and find out how microchips, computers, telecom and the Internet came into being. Who were the key players and what were their key contributions? What were the underlying concepts in this complex set of technologies? What is the digital technology that is leading to the convergence of computers, communication, media, movies, music and education? Who have been the Indian scientists and technologists who played a significant role in this global saga and what did they actually do?

Without being parochial, it is important to publicise the Indian contribution to IT for its inspirational role for youth.

In the last two decades, we have seen some attitudinal changes too. The fear that computerisation will lead to mass unemployment has vanished. We have witnessed old jobs being done with new technology and new skills, and with the added bonus of efficiency and convenience. The transformation brought about at the reservation counters of the Indian Railways and in bank branches are examples of this. Moreover, several hundred thousand jobs have been created in the IT sector for software programmers and hardware engineers.

Today, we have a vibrant software services industry, built in the last thirty years by Indian entrepreneurs, which is computerising the rest of the world. Indian IT professionals have built a reputation all over the world as diligent problem solvers and as lateral thinkers. Hundreds of Indian engineers have not only contributed to the development of innovative technology but also succeeded as entrepreneurs in the most competitive environments. As R.A. Mashelkar says, “It is the convergence of Laxmi and Saraswati.”

A globalising world is discovering that world-class services can be provided by Indian accountants, financial experts, bankers, doctors, architects, designers, R&D scientists et at. Thanks to the development of modern telecom infrastructure, they can provide it without emigrating from India. In the seventies and eighties many of us used to lament that India had missed the electronics bus. Today, however, due to the development of skills in microchip design, engineers in India are designing cutting-edge chips, and communication software engineers are enabling state-of-the-art mobile phones and satellite phones.

While all this is laudable, it also begs a question: how will IT impact ordinary Indians? Can a farmer in Bareilly or Tanjavur; or a student in a municipal school in Mumbai; or a sick person in a rural health centre in hilly Garhwal; or an Adivasi child in Jhabua or Jharkhand benefit from an IT-enabled Indian nation? I believe they can, and must.

In this book, I have attempted to espouse a complex set of technologies in relatively simple terms. Stories and anecdotes have been recounted to give a flavour of the excitement. A bibliography is presented at the end of each chapter for the more adventurous reader, along with website addresses, where available.

****

OF CHIPS AND WAFERS

"The complexity [of integrated circuits] for minimum costs has increased at a rate of roughly a factor of two per year."
— **GORDON E MOORE**,
Electronics, VOL 38, NO 8, 1965

Where Silicon and Carbon atoms will
Link valencies, four figured, hand in hand
With common Ions and Rare Earths to fill
The lattices of Matter, Glass or Sand,
With tiny Excitations, quantitatively grand

— FROM "*The Dance of the Solids*", BY **JOHN UPDIKE**
(*Midpoint and Other Poems,* ALFRED A KNOPF, 1969)

Several technologies and theories have converged to make modern Information Technology possible. Nevertheless, if we were to choose one that has laid the ground for revolutionary changes in this field, then it has to be semiconductors and microelectronics. Complex electronic circuits made of several components integrated on a single tiny chip of silicon are called Integrated Circuits or chips. They are products of modern microelectronics.

Chips have led to high-speed but inexpensive electronics. They have broken the speed, size and cost barriers and made electronics available to millions of people. This has created discontinuities in our lives—in the way we communicate, compute and transact.

The chip industry has created an unprecedented disruptive technology that has led to falling prices and increasing functionality at a furious pace.

DECONSTRUCTING MOORE'S LAW

Gordon Moore, the co-founder of Intel, made a prediction in 1965 that the number of transistors on a chip and the raw computing power of microchips would double every year while the cost of production would remain the same. When he made this prediction, chips had only 50 transistors; today, a chip can have more than 250 million transistors. Thus, the power of the chip has increased by a factor of five million in about thirty-eight years. The only correction to Moore's Law is that nowadays the doubling is occurring every eighteen months, instead of a year.

As for cost, when transistors were commercialised in the early 1950s, one of them used to be sold for $49.95; today a chip like Pentium-4, which has 55 million transistors, costs about $200. In other words, the cost per transistor has dropped by a factor of ten million.

This is what has made chips affordable for all kinds of applications: personal computers that can do millions of arithmetic sums in a second, telecom networks that carry billions of calls, and Internet routers that serve up terabytes of data (tera is a thousand billion). The reduced costs allow chips to be used in a wide range of modern products. They control cars, microwave ovens, washing machines, cell phones, TVs, machine tools, wrist-watches, radios, audio systems and even toys. The Government of India is toying with the idea of providing all Indians with a chip-embedded identity card carrying all personal data needed for public purposes.

According to the Semiconductor Industry Association of the US, the industry is producing 100 million transistors per year for every person on earth (6 billion inhabitants), and this figure will reach a billion transistors per person by 2008!

The semiconductor industry is estimated to be a $300 billion-a-year business. Electronics, a technology that was born at the beginning of the twentieth century, has today been integrated into everything imaginable. The Nobel Committee paid the highest tribute to this phenomenal innovation in the year 2000 when it awarded the Nobel Prize in physics to Jack Kilby, who invented the integrated circuit, or the chip, at Texas Instruments in 1958.

Considering the breathtaking advances in the power of chips and the equally astonishing reduction in their cost, people sometimes wonder whether this trend will continue forever. Or will the growth come to an end soon?

The Institute of Electrical and Electronics Engineers, or (IEEE as 'I-triple E')—the world's most prestigious and largest professional association of electrical, electronics and computer engineers, conducted a survey among 565 of its distinguished fellows, all highly respected technologists. One of the questions the experts were asked was: how long will the semiconductor industry see exponential growth, or follow Moore's Law? The results of the survey, published in the January 2003 issue of *IEEE Spectrum* magazine, saw the respondents deeply divided. An optimistic seventeen per cent said more than ten years, a majority—fifty two per cent—said five to ten years and a pessimistic thirty per cent said less than five years. So much for a 'law'!

Well, then, what has fuelled the electronics revolution? The answer lies in the developments that have taken place in semiconductor physics and microelectronics. Let us take a quick tour of the main ideas involved in them.

ALL ABOUT SEMICONDUCTORS

What are semiconductors? A wit remarked, "They are bus conductors who take your money and do not issue tickets." Jokes apart, they are materials that exhibit strange electrical properties. Normally, one comes across metals like copper and aluminium, which are good conductors, and rubber and wood, which are insulators, which do not conduct electricity. Semiconductors lie between these two categories.

What makes semiconductors unique is their behaviour when heated. All metals conduct well when they are cold, but their conductivity decreases when they become hot. Semiconductors do the exact opposite: they become insulators when they are cold and mild conductors when they are hot. So what's the big deal? Well, classical nineteenth century physics, with its theory of how materials conduct or insulate the flow of electrons—tiny, negatively charged particles—could not explain this abnormal behaviour. As the new quantum theory of matter evolved in 1925-30, it became clear why semiconductors behave the way they do.

Quantum theory explained that, in a solid, electrons could have energies in two broad ranges: the valence band and the conduction band. The latter is at a higher level and separated from valence band by a gap in energy known as the band gap. Electrons in the valence band are bound to the positive part of matter and the ones in the conduction band are almost free to move around. For example, in metals, while some electrons are bound, many are free. So metals are good conductors.

According to atomic physics, heat is nothing but energy dissipated in the form of the random jiggling of atoms. At lower temperatures, the atoms are relatively quiet, while at higher temperatures they jiggle like mad. However, this jiggling slows down the motion of electrons through the material since they get scattered by jiggling atoms. It is similar to a situation where you are trying to get through a crowded hall. If the people in the crowd are restive and randomly moving then it takes longer for you to move across than when they are still. That is the reason metals conduct well when they are cold and conduct less as they become hotter and the jiggling of the atoms increases.

In the case of semiconductors, there are no free electrons at normal temperatures, since they are all sunk into the valence band, but, as the temperature increases, the electrons pick up energy from the jiggling atoms and get kicked across the band gap into the conduction band. This new-found freedom of a few electrons makes the semiconductors mild conductors at higher temperatures. To increase or decrease this band gap, to shape it across the length of the material the way you want, is at the heart of semiconductor technology.

Germanium, an element discovered by German scientists and named after their fatherland, is a semiconductor. It was studied extensively. When the UK and the US were working on a radar project during the Second World War, they heavily funded semiconductor research to build new electronic devices. Ironically, the material that came to their assistance in building the radar and defeating Germany was germanium.

MISCHIEF OF THE MISFITS

Now, what if small amounts of impurities are introduced into semiconductors? Common sense says this should lead to small changes in their properties. But, at the atomic level, reality often defies commonsense. Robert Pohl, who pioneered experimental research into semiconductors, noticed in the 1930s that the properties of semiconductors change drastically if small amounts of impurities are added to the crystal.

This was the outstanding feature of these experiments and what Nobel laureate Wolfgang Pauli called 'dirt physics'. Terrible as that sounds, the discovery of this phenomenon later led to wonderful devices like diodes and transistors. The 'dirty' semiconductors hit pay dirt.

Today, the processes of preparing a semiconductor crystal are advanced and the exact amount of a particular impurity to be added to it is carefully controlled in parts per million. The process of adding these impurities is called 'doping'.

If we experiment with silicon, which has four valence electrons, and dope it with minuscule amounts (of the order of one part in a million) of phosphorus, arsenic or antimony, boron, aluminium, gallium or indium, we will see the conductivity of silicon improve dramatically.

How does doping change the behaviour of semiconductors drastically? We can call it the mischief of the misfits. Misfits, in any ordered organisation, are avoided or looked upon with deep suspicion. But there are two kinds of misfits: those that corrupt and disorient the environment are called 'bad apples'; those that stand above the mediocrity around them, and might even uplift the environment by seeding it with change for the better, are called change agents. The proper doping of pure, well-ordered semiconductor crystals of silicon and germanium leads to dramatic and positive changes in their electrical behaviour. These 'dopants' are change agents.

How do dopants work? Atomic physics has an explanation. Phosphorus, arsenic and antimony all have five electrons in the highest energy levels. When these elements are introduced as impurities in a silicon crystal and occupy the place of a small number of silicon atoms in a crystal, the crystal structure does not change much. But, since the surrounding silicon atoms have four electrons each, the extra electron in each dopant, which is relatively unattached, gets easily excited into the conduction band at room temperature. Such doped semiconductors are called N-type (negative type) semiconductors. The doping materials are called 'donors'.

On the other hand, when we use boron, aluminium, gallium or indium as dopants, they leave a gap, or a 'hole', in the electronic borrowing and lending mechanisms of neighbouring atoms in the crystal, because they have three valence electrons. These holes, or deficiency of electrons, act like positively charged particles. Such semiconductors are described as Ptype (positive type). The dopants in this case are called 'acceptors'.

VALVES, TRANSISTORS, *et al*

In the first four decades of the twentieth century, electronics was symbolized by valves. Vacuum tubes, or valves, which looked like dim incandescent light bulbs, brought tremendous change in technology and made radio and TV possible. They were the heart of both the transmission stations and the receiving sets at home, but they suffered from some big drawbacks: they consumed a lot of power, took time to warm up and, like ordinary light bulbs, burnt out often and unpredictably. Thus, electronics faced stagnation.

The times were crying for a small, low-power, low-cost, reliable replacement for vacuum tubes or valves. The need became all the more urgent with the development of radar during the Second World War.

Radars led to the development of microwave engineering. A vacuum tube called the magnetron was developed to produce microwaves. What was lacking was an efficient detector of the waves reflected by enemy aircraft. If enemy aircraft could be detected as they approached a country or a city, then precautionary measures like evacuation could minimise the damage to human life and warn the anti-aircraft guns to be ready. Though it was a defensive system, the side that possessed radars suffered the least when airpower was equal, and hence it had the potential to win the war.

This paved the way for investments in semiconductor research, which led to the development of semiconductor diodes.

It is estimated that more money was spent on developing the radar than the Manhattan Project that created the atom bomb. Winston Churchill attributed the allied victory in the air war substantially to the development of radar.

Actually, electronics hobbyists knew semiconductor diodes long ago. Perhaps people in their middle age still remember their teenage days when crystal radios were a rage. Crystals of galena (lead sulphide), with metal wires pressed into them and called 'cat's whiskers', were used to build inexpensive radio sets. It was a semiconductor device. The crystal diode converted the incoming undulating AC radio waves into a unidirectional DC current, a process known as 'rectification'. The output of the crystal was then fed into an earphone.

A rectifier or a diode is like a one-way valve used by plumbers, which allows water to flow in one direction but prevents it from flowing back.

Interestingly, Indian scientist Jagdish Chandra Bose, who experimented with electromagnetic waves during the 1890s in Kolkata, created a semiconductor microwave detector, which he called the 'coherer'. It is believed that Bose's coherer, made of an iron-mercury compound, was the first solid-state device to be used. He demonstrated it to the Royal Institution in London in 1897. Guglielmo Marconi used a version of the coherer in his first wireless radio in 1897.

Bose also demonstrated the use of galena crystals for building receivers for short wavelength radio waves and for white and ultraviolet light. He received patent rights, in 1904, for their use in detecting electromagnetic radiation. Neville Mott, who was awarded the Nobel Prize in 1977 for his contributions to solid-state electronics, remarked, "J.C. Bose was at least 60 years ahead of his time" and "In fact, he had anticipated the existence of P-type and N-type semiconductors."

Semiconductor diodes were a good beginning, but what was actually needed was a device that could amplify signals. A 'triode valve' could do this but had all the drawbacks of valve technology, which we referred to earlier. The question was: could the semiconductor equivalent of a triode be built?

For a telephone company, a reliable, inexpensive and low-power consuming amplifier was crucial for building a long-distance communications network, since long-distance communications are not possible without periodic amplification of signals. This led to AT&T, which had an excellent research and development laboratory named after Graham Bell, called Bell Labs, in New Jersey, starting a well-directed effort to invent a semiconductor amplifier.

William Shockley headed the Bell Labs research team. The team consisted, among others, of John Bardeen and Walter Brattain. The duo built an amplifier using a tiny germanium crystal. Announcing the breakthrough to a yawning bunch of journalists on 30 June 1948, Bell Labs' Ralph Bown said: "We have called it the transistor because it is a resistor or semiconductor device which can amplify electrical signals as they are transferred through it."

The press hardly took note. A sympathetic journalist wrote that the transistor *might* have some applications in making hearing aids! With apologies to T S Eliot, thus began the age of solid-state electronics—"not with a bang, but a whimper".

The original transistor had manufacturing problems. Besides, nobody really understood how it worked. It was put together by tapping two wires into a block of germanium. Only some technicians had the magic touch that made it work. Shockley ironed out the problems by creating the junction transistor in 1950, using junctions of N-type and P-type semiconductors.

SAND CASTLES OF A DIFFERENT KIND

The early transistors, which were germanium devices, had a problem. Though germanium was easy to purify and deal with, devices made from it had a narrow temperature range of operation. Thus, if they heated up beyond sixty-seventy degrees centigrade, they behaved erratically. So the US military encouraged research into materials that would be more robust in battlefield conditions (rather than laboratories and homes).

A natural choice was silicon. It did not have some of the good properties of germanium. It was not easy to prepare pure silicon crystals, but silicon could deliver good results over a wide range of temperatures, up to 200 degrees centigrade. Moreover, it was easily available. Silicon is the second most abundant element on earth, constituting twenty-seven per cent of the earth's crust. Ordinary sand is an oxide of silicon.

In 1954, Texas Instruments commercialised the silicon transistor and tried marketing a portable radio made from it. It was not so successful, but a fledgling company in post-war Japan, called Sony, was. Portable radios became very popular and, for many years and for most people, the word transistor became synonymous with an inexpensive portable radio.

What makes a transistor such a marvel? To understand a junction transistor, imagine a smooth road with a speed breaker. Varying the height of the speed breaker controls the traffic flow. However the effect of the change in the height of the 'potential barrier' in the transistor's sandwiched region, which acts like a quantum speed breaker on the current, is exponential. That is, halving the height of the barrier or doubling it does not halve or double the current. Instead, it cuts it down to a seventh of its value or increases it seven times, thereby providing the ground for the amplification effect. After all, what is amplification but a small change getting converted to a large change? Thus, a small electrical signal can be applied to the 'base' of the transistor to lead to large changes in the current between the 'emitter' and the 'collector'.

FRETTING OVER FETS

Then came the 'FET'. The idea was to take a piece of germanium, doped appropriately, and directly control the current by applying an electric field across the flow path through a metal contact, fittingly called a gate. This would be a 'field effect transistor', or FET.

While Bell Labs' Bardeen and Brattain produced the transistor, their team leader, Shockley, followed a different line; he was trying to invent the FET. Bardeen and Brattain beat him to inventing the transistor, and the flamboyant Shockley could never forget that his efforts failed while his team members' approach worked. This disappointment left its mark on an otherwise brilliant career. Shockley's initial effort did not succeed because the gate started drawing current. Putting an insulator between the metal and the semiconductor was a logical step, but efforts in this direction failed until researchers abandoned their favourite germanium for silicon.

We have already mentioned the better temperature range of silicon. But silicon had one major handicap: as soon as pure silicon was exposed to oxygen it 'rusted' and a highly insulating layer of silicon dioxide was formed on the surface. Researchers were frustrated by this silicon rusting.

Now that a layer of insulating material was needed between the gate and the semiconductor for making good FETs, and germanium did not generate insulating rust, silicon, which developed insulating rust as soon as it was exposed to oxygen, became a natural choice. Thus was born the 'metal oxide semiconductor field effect transistor', or MOSFET. It is useful to remember this rather long acronym, since MOSFETs dominate the field of microelectronics today.

A type of MOSFET transistor called CMOS (complementary metal oxide semiconductor) was invented later. This had the great advantage of not only operating at low voltages but also dissipating the lowest amount of heat. A large number of CMOS transistors can be packed per square inch, depending on how sharp is the 'knife' used to cut super-thin grooves on thin wafers of silicon. *Today CMOS is the preferred technology in all microchips.*

INVENTION OF THE IC

The US military was pushing for the micro-miniaturisation of electronics. In 1958, Texas Instruments hired Jack Kilby, a young PhD, to work on a project funded by the US defence department. Kilby was asked if he could do something about a problem known as the 'tyranny of numbers'. It was a wild shot. Nobody believed that the young man would solve it.

What was this 'tyranny of numbers', a population explosion? Yes, but of a different kind. As the number of electronic components increased in a system, the number of connecting wires and solders also increased. The fate of the whole system not only depended on whether every component worked but also whether every solder worked. Kilby began the search for a solution to this problem.

Americans, whether they are in industry or academia, have a tradition of taking a couple of weeks' vacation during summer. In the summer of 1958, Kilby, who was a newcomer to his assignment, did not get his vacation and was left alone in his lab while everyone else went on holiday. The empty lab gave Kilby an opportunity to try out fresh ideas.

"I realised that semiconductors were all that were really required. The resistors and capacitors could be made from silicon, while germanium was used for transistors," Kilby wrote in a 1976 article titled *Invention of the IC*. "My colleagues were skeptical and asked for some proof that circuits made entirely of semiconductors would work. I therefore built up a circuit using discrete silicon elements. By September, I was ready to demonstrate a working integrated circuit built on a piece of semiconductor material."

Several executives, including former Texas Instruments chairman Mark Shepherd, gathered for the event on 12 September 1958. What they saw was a sliver of germanium, with protruding wires, glued to a glass slide.

It was a rough device, but when Kilby pressed the switch the device showed clear amplification with no distortion. His invention worked. He had solved the problem—and he had invented the integrated circuit.

Did Kilby realise the significance of his achievement? "I thought it would be important for electronics as we knew it then, but that was a much simpler business," said Kilby when the author interviewed him in October 2000 in Dallas, Texas, soon after the announcement of his Nobel Prize award. "Electronics was mostly radio and television and the first computers. What we did not appreciate was how lower costs would expand the field of electronics beyond imagination. It still surprises me today. The real story has been in the cost reduction, which has been much greater than anyone could have anticipated."

The unassuming Kilby was a typical engineer who wanted to solve problems. In his own words, his interest in electronics was kindled when he was a kid growing up in Kansas. "My dad was running a small power company scattered across the western part of Kansas. There was this big ice storm that took down all the telephones and many of the power lines, so he began to work with amateur radio operators to provide some communications. That was the beginning of my interest in electronics."

His colleagues at Texas Instruments challenged Kilby to find a use for his integrated circuits and suggested that he work on an electronic calculator to replace large mechanical ones. This led to the successful invention of the electronic calculator. In the 1970s calculators made by Texas Instruments were a prized possession among engineering students. In a short period of time the electronic calculator replaced the old slide rule in all scientific and engineering institutions. It can truly be called the first mass consumer product of integrated electronics.

Meanwhile, Shockley, the co-inventor of the transistor, had walked out of Bell Labs to start Shockley Semiconductor Laboratories in California. He assembled a team consisting of Robert Noyce, Gordon Moore and others. However, though Shockley was a brilliant scientist, he was a poor manager of men. Within a year, a team of eight scientists led by Noyce and Moore left Shockley Semiconductors to start a semiconductor division for Fairchild Camera Inc.

Said Moore, "We had a few other ideas coming along at that time. One of them was something called a planar transistor, created by Jean Hoerni, a Caltech post-doc. Jean was a theoretician, and so was not very useful when we were building furnaces and all that kind of stuff. He just sat in his office, scribbling things on a piece of paper, and he came up with this idea for building a transistor by growing a silicon oxide layer over the junctions. Nobody had ever tried leaving the oxide on. When we finally got around to trying it, it turned out to be a great idea; it solved all the previous surface problems. Then we wondered what else we might do with this planar technology. Robert Noyce came up with the two key inventions to make a practical integrated circuit: by leaving the oxide on, one could run interconnections as metal films over the top of its devices; and one could also put structures inside the silicon that isolated one transistor from the other."

While Kilby's invention had individual circuit elements connected together with gold wires, making the circuit difficult to scale up, Hoerni and Noyce's planar technology set the stage for complex integrated circuits. Their ideas are still the basis of the process used today. Though Kilby got the Nobel Prize, Noyce and Kilby share the credit of coming up with the crucial innovations that made an integrated circuit possible.

After successfully developing the IC business at Fairchild Semiconductors, Noyce and Moore were again bit by the entrepreneurial bug. In 1968 they seeded a new company, Intel, which stood for Integrated Electronics. Intel applied the IC technology to manufacture semiconductor-based memory and then invented the microprocessor. These two concepts have powered the personal computer revolution of the last two decades.

In Kilby and Noyce's days, one could experiment easily with IC technology. "No equipment cost more than $10,000 during those days," says Kilby. Today chip fabrication plants, called 'Fabs', cost as much as two to three billion dollars.

Let us look at the main steps involved in fabricating a chip today in a company like Intel. If you are a cooking enthusiast then it might remind you of a layered cake. Craig Barret, explained the process in an article in 1998: *'From Sand to Silicon: Manufacturing an Integrated Circuit'*.

'PRINTING' CHIPS

The chip-making process, in its essence, resembles the screen-printing process used in the textile industry. When you have a complicated, multi coloured design to be printed on a fabric, the screen printer takes a picture of the original, transfers it to different silk screens by a photographic process, and then uses each screen as a stencil while the dye is rolled over the screen. One screen is used for each colour. The only difference is in the size of the design. With dress material, print sizes run into square metres; with chips, containing millions of transistors (the Pentium-4, for example, has fifty-five million transistors), each transistor occupies barely a square micron. How is such miniature design achieved? There are all kinds of superfine works of art, including calligraphy of a few words on a grain of rice. But the same grain of rice can accommodate a complicated circuit containing about 3,000 transistors! How do chipmakers pull off something so incredible?

In a way, the chip etcher's approach is not too different from that of the calligraphist writing on a grain of rice. While the super-skilled calligraphist uses an ordinary watchmaker's eyepiece as a magnifying glass, the chipmaker uses very short wavelength light (ultraviolet light) and sophisticated optics to reduce the detailed circuit diagrams to a thousandth of their size. These films are used to create stencils (masks) made of materials that are opaque to light.

The masks are then used to cast shadows on photosensitive coatings on the silicon wafer, using further miniaturisation with the help of laser light, electron beams and ultra-sophisticated optics to imprint the circuit pattern on the wafer.

The process is similar to the good old printing technology called lithography, where the negative image of a text or graphic is transferred to a plate covered with photosensitive material, which is then coated by ink that is transferred to paper pressed against the plates by rollers. This explains why the process of printing a circuit on silicon is called photolithography.

Of course, we are greatly simplifying the chip-making methodology for the sake of explaining the main ideas. In actual fact, several layers of materials—semiconductors and metals—have to be overlaid on each other, with appropriate insulation separating them. Chipmakers use several sets of masks, just as newspaper or textile printers use different screens to imprint different colours in varied patterns.

While ordinary printing transfers flat images on paper or fabric, chipmakers create three-dimensional structures of micro hills and vales by using a host of chemicals for etching the surface of the silicon wafer.

The fineness of this process is measured by how thin a channel you can etch on silicon. So, when someone tells you about 0.09-micron technology being used by leading chipmakers, they are referring to hitech scalpels that can etch channels as thin as 0.09 micron. To get a sense of proportion, that is equivalent to etching 350 parallel ridges and vales on a single strand of human hair!

Only a couple of years ago, most fabs used 0.13-micron technology; today, many leading fabs have commercialised 0.09-micron technology and are experimenting with 0.065-micron technology in their labs.

What does this mean? Well, roughly each new technology is able to etch a transistor in half the surface area of the silicon wafer than the previous one. Lo and behold, the “secret” of Moore’s Law of doubling transistor density on a chip!

WHY MOORE’S LAW MUST END

What are the problems in continuing this process? Making the scalpels sharper is one. Sharper scalpels mean using shorter and shorter wavelengths of light for etching. But, as the wavelength shortens we reach the X-ray band, and we do not yet have X-ray lasers or optics of good quality in that region.

There is another hurdle. As circuit designs get more complex and etching gets thinner, the masks too become thinner. A law in optics says that if the dimensions of the channels in a mask are of the order of the wavelength of light, then, instead of casting clear shadows, the masks will start ‘diffracting’—bands of bright and dark regions would be created around the edges of the shadow, thereby limiting the production of sharply defined circuits.

Moreover, as the channels get thinner there are greater chances of electrons from one channel crossing over to the other due to defects, leading to a large number of chips failing at the manufacturing stage.

Surprisingly, though, ingenious engineers have overcome the hurdles and come up with solutions that have resulted in further miniaturisation. Until now Moore’s Law has remained a self-fulfilling prophecy.

EXTENDING THE TENURE OF MOORE’S LAW

What has been achieved so far has been extraordinary. But it has not been easy. At every stage, engineers have had to fine-tune various elements of the manufacturing process and the chips themselves.

For example, in the late 1970s, when memory chipmakers faced the problem of limited availability of surface, they found an innovative answer to the problem. “The dilemma was,” says Pallab Chatterjee, “should we build skyscrapers or should we dig underground into the substrate and build basements and subways?”

While working at Texas Instruments in the 1970s and 1980s, Chatterjee played a major role in developing reliable micro transistors and developing the ‘trenching’ technology for packing more and more of them per square centimetre. This deep sub-micron technology resulted in the capacity of memory chips leapfrogging from kilobytes to megabytes. Texas Instruments was the first to introduce a 4 MB DRAM memory, back in 1985. Today, when we can buy 128 MB or 256 MB memory chips in any electronics marketplace for a few thousand rupees, this may seem trite; but the first 4 MB DRAM marked a big advance in miniaturisation.

Another person of Indian origin, Tom Kailath, a professor of communication engineering and information theory at Stanford University in the US, developed signal processing techniques to compensate for the diffractive effects of masks. A new company, Numerical Technologies, has successfully commercialised Kailath’s ideas. Kailath’s contribution was an instance of the cross-fertilisation of technologies, with ideas from one field being applied to solve problems in a totally different field. Well known as a leading academic and teacher, Kailath takes great satisfaction in seeing some of his highly mathematical ideas getting commercialized in a manufacturing environment.

Another leading researcher in semiconductor technology who has contributed to improving efficiencies is Krishna Saraswat, also at Stanford University. "When we were faced with intense competition from Japanese chipmakers in the 1980s, the Defence Advanced Research Projects Agency (DARPA), a leading financer of hi-tech projects in the US, undertook an initiative to improve fabrication efficiencies in the American semiconductor industry," says Chatterjee. "We at Texas Instruments collaborated with Saraswat at Stanford, and the team solved the problems of efficient batch processing of silicon wafers."

HIGH-COST BARRIERS

One of the ways diligent Japanese companies became more efficient than the Americans was by paying attention to 'clean-room' conditions. Chatterjee and Saraswat spotted it and brought about changes in manufacturing techniques that made the whole US chip industry competitive. One of Saraswat's main concerns today is to reduce the time taken by signals to travel between chips and even within chips. "The 'interconnects' between chips can become the limiting factor to chip speeds, even before problems are faced at the nano-physics level," he explains.

Every step of the chip-manufacturing process has to be conducted in ultra dust-free clean rooms; every gas or chemical used—including water and the impurities used for doping—have to be ultra pure! When the author visited the Kilby Centre (a state-of-the-art R&D centre set up by Texas Instruments and named after its most famous inventor) at Dallas in the year 2000, they were experimenting with 0.90-micron technology. The technicians inside the clean rooms resembled astronauts in spacesuits.

All this translates into the high capital costs of chip fabrication facilities today. In the 1960s it cost a couple of million dollars to set up a fab; today it costs a thousand times more. The high cost of the fabs creates entry barriers to newcomers in microelectronics. Besides, chip making is still an art and not really a science. Semiconductor companies use secret recipes and procedures much like gourmet cooking. Even today, extracting the maximum from a fab is the key to success in semiconductor manufacturing.

If the capital costs are so high, how are chips getting cheaper? The answer lies in volumes. A new fab might cost, say, five billion dollars, but if it doubles the number of transistors on a chip and produces chips in the hundreds of millions, then the additional cost per chip is marginal, even insignificant. Having produced high-performance chips with new technology, the manufacturer also receives an extra margin on each chip for a year or so and recovers most of its R&D and capital costs. After that the company can continue to fine-tune the plant, while reducing the price, and still remain profitable on thin margins.

THE ENTRAILS OF A CHIP

Though the transistor was invented to build an amplifier, the primary use of the transistor in a chip today is as a switch—a device that conducts or does not conduct, depending on the voltage applied to the gate. The 'on' state represents a 1 and the 'off' state represents a 0, and we have the basic building block of digital electronics. These elements are then used to design logic gates.

What are logic gates? They are not very different from ordinary gates, which let people pass through if they have the requisite credentials. A fundamental gate from which all other logic gates can be built is called a NAND gate. It compares two binary digital inputs, which can be either 1 or 0.

If the values of both inputs are 1, then the output value is 0; but if the value of one input is 0 and that of the other is 1, or if the values of both inputs are 0, the output value is 1.

These gates can be configured to carry out higher-level functions. Today chips are designed with millions of such gates to carry out complex functions such as microprocessors in computers or digital signal processors in cell phones.

Simpler chips are used in everyday appliances. Called microcontrollers, they carry out simple functions like directing the electronic fuel injection system in your car, adjusting contrast, brightness and volume in your TV set, or starting different parts of the wash cycle at the right time in your washing machine.

"Earlier, there used to be audio amplifiers with four transistors; today even a simple audio chip has 2,000 transistors," says Sorab Ghandhi, who, in 1953, wrote the first-ever book on transistor circuit design.

DID INDIA MISS THE MICROCHIP BUS?

Vinod Dham, who joined Intel in the mid-1970s and later led the project that created the Pentium, the most successful Intel chip to date, has an interesting story to tell. He says: "Gurpreet Singh, who, back in the sixties, founded Continental Devices—one of the first semiconductor companies in India and the place where I cut my teeth in the early seventies—told me that Bob Noyce came and stayed with him in Delhi in the sixties. Noyce spent fifteen days trying to convince the Indian government to allow Intel to establish a chip company in India!"

The Indian government rejected the proposal. Why did it adopt such an attitude towards electronics and computers in general? It seems inexplicable.

There are many horror stories told by industry veterans about how many times India missed the bus. According to Bishnu Pradhan, who led the R&D centre at Tata Electric Companies for two decades and later led C-DOT (Centre for Development of Telematics), prototypes of personal computers were being made in India way back in the 1970s. These PCs were as sophisticated as those being developed in the Silicon Valley. But the Indian government discouraged these attempts on one pretext or another. That is why, while India has supplied chip technologists to other countries, several countries, which were way behind India in the 1960s, are today leagues ahead of us. Taiwan and South Korea are two such examples.

Even the much touted software industry in India had to struggle due to the lack of computers. People like F.C. Kohli, who led Tata Consultancy Services for three decades, had to spend a lot of time and effort convincing the government to allow the import of computers to develop software.

In the case of nuclear and space technologies, Homi Bhabha, Vikram Sarabhai and Satish Dhawan fully utilised foreign assistance, know-how and training to catch up with the rest of the world. Only when other countries denied these technologies to them did they invest R&D resources in developing them indigenously. They were not dogmatic; they were global in outlook and cared for national interests as well. Unfortunately, India missed that kind of leadership in policy-making in electronics and computers.

After much confabulation, the Indian government bought a fab in the 1980s and established the Semiconductor Complex Ltd at Chandigarh. But the facility was burnt down in a fire in the mid-eighties. It has since been rebuilt, but it was too little too late. SCL's technology remains at the one-micron level while the world has moved to 0.09 micron.

A modern fab in the country would have given a boost to Indian chip designers; they could not only have designed chips but also tested their innovative designs by manufacturing in small volumes. The fab could have accommodated such experiments while doing other, high-volume work for its regular business. Today SCL has opened its doors for such projects but, according to many experts, it is uncompetitive.

SOFTENING OF THE HARDWARE

If India is uncompetitive in this business, how should one interpret newspaper reports about young engineers in Bangalore and Pune designing cutting-edge chips? How has that happened?

This has been made possible by another major development in semiconductor technology: separation of the hardware from the software. What does this mean? That you can have somebody designing a chip in some place on his workstation—a powerful desktop computer—and get it fabricated elsewhere. There is a separation of chip design and fabrication. As a result, there are fabs that just fabricate chips, and there are 'fabless chip companies' which only design chips. Some enthusiasts call them 'fabulous chip companies'.

It is not very different from the separation that took place long ago between the civil engineers who build houses and the architects who design them. If we go a step further and devise programmes to convert the ideas of architects into drawings on the computer, they are called 'computer aided design', or CAD, packages.

Interestingly, in 1980, when Vinod Khosla, a twenty-five-year-old engineer, started a CAD software company, Daisy Systems, to help in chip design, he found that such software needed powerful workstations, which did not then exist. That led to Khosla joining Andreas Bechtolsheim, Bill Joy and Scott McNealy to co-found Sun Microsystems in the spring of 1982.

Khosla recalls, "When I was fifteen-sixteen and living in Delhi, I read about Intel, a company started by a couple of PhDs. Those days I used to go to Shankar Market and rent old issues of electronics trade journals in order to follow developments. Starting a hi-tech business was my dream long before I went to the Indian Institute of Technology in Delhi. In 1975, even before I finished my B.Tech, I tried to start a company. But in those days you couldn't do this in India if your father did not have 'connections'. That's why I resonate with role models. Bob Noyce, Gordon Moore and Andy Grove at Intel became role models for me."

Today Sun is a broad-based computer company. Khosla was the chief executive of Sun when he left the company in 1985 and became a venture capitalist. Today he is a partner in Kleiner Perkins Caulfield Byers and is voted, year on year, with boring repetition, as a top-notch venture capitalist in Silicon Valley. Meanwhile, Sun workstations continue to dominate chip design.

CAD is only a drawing tool that automates the draughtsman's work. How do you convert the picture of a transistor into a real transistor on silicon? How do you pack a lot of transistors on the chip without them overlapping or interfering with each other's function? Can you go up the ladder of abstraction and convert the logical operations expressed in Boolean equations into transistor circuits? Can you take one more step and give the behaviour of a module in your circuitry and ask the tool to convert that into a circuit?

Designing a circuit from scratch, using the principles of circuit design, would take a lot of time and money. There would be too many errors, and each designer would have his own philosophy, which might not be transparent to the next one who wished to debug it. Today's tools can design circuits if you tell them what functionality you want. Which means that if you write down your specifications in a higher-level language, the tools will convert them into circuits.

What sounded like a wish list from an electronics engineer has become a reality in the last forty years, thanks to electronic design automation, or EDA, tools. The trend to develop such tools started in the 1960s and '70s but largely remained the proprietary technology of chipmakers. Yet, thanks to EDA tools, today's hardware designers use methods similar to those that software designers use—they write programs and let tools generate the implementation. Special languages known as hardware description languages have been developed to do this. That is the secret behind designers in Bangalore and Pune developing cutting-edge chips.

In a sense, India is catching the missed electronics bus at a different place, one called chip design.

Interestingly, several Indians have played a pioneering role in developing design tools. Raj Singh, a chip designer who co-authored one of the earliest and most popular books on hardware description languages, and later went on to build several start-ups, talks of Suhas Patil. "Suhas had set up Patil Systems Inc. as a chip-design company in Utah based upon his research in Storage Logic Arrays at the Massachusetts Institute of Technology," says Singh. "He moved it later to the Silicon Valley as SLA Systems to sell IC design tools. Finding it difficult to sell tools, he changed the business to customer-specific ICs using his SLA toolkit and founded Cirrus Logic as a fabless semiconductor company."

Verilog, a powerful hardware description language, was a product of Gateway Automation, founded by Prabhu Goel in Boston. Goel had worked on EDA tools at IBM from 1973-82 and then left IBM to start Gateway. Goel's Gateway was also one of the first companies to establish its development centre in India.

BANGALORE BLOOMS

The first multinational company to establish a development centre in India was the well-known chip company Texas Instruments, which built a facility in Bangalore in 1984. The company's engineers in Bangalore managed to communicate directly with TI in Dallas via a direct satellite link—another first. This was India's first brush with hi-tech chip design.

"Today TI, Bangalore, clearly is at the core of our worldwide network and has proved that cutting-edge work can be done in India," says K. Bala, chief operating officer at TI, Japan, who was earlier in charge of the Kilby Centre in Dallas. "We have produced over 200 patents and over 100 products for Texas Instruments in the last five years with a staff that constitutes just two per cent of our global workforce," says a proud Bobby Mitra, the managing director of the company's Indian operations.

The success of Texas Instruments has not only convinced many other multinational companies like Analog Devices, National Semiconductor and Intel to build large chip-designing centres in India, it has also led to the establishment of Indian chip design companies. "Indian technologists like Vishwani Agarwal of Bell Labs have helped bring international exposure to Indian chip designers by organising regular international conferences on VLSI design in India," says Juzer Vasi of IIT, Bombay, which has become a leading educational centre for microelectronics.

DESIGNS ON DESIGN

Where are we heading next from the design point of view? "Each new generation of microprocessors that is developed using old design tools leads to new and more powerful workstations, which can design more complex chips, and hence the inherent exponential nature of growth in chip complexity," says Goel.

"The next big thing will be the programmable chip," says Suhas Patil. Today if you want to develop a chip that can be used for a special purpose in modest numbers, the cost is prohibitive. The cost of a chip comes down drastically only when it is manufactured in the millions. Patil hopes that the advent of programmable chips will allow the design of any kind of circuit on it by just writing a programme in C language. "Electronics will become a playground for bright software programmers, who are in abundant numbers in India, but who may not know a thing about circuits," says Patil. "This will lead to even more contributions from India."

There is another aspect of chip making and it's called testing and verification. How do you test and verify that the chip will do what it has been designed to? "Testing a chip can add about fifty per cent to the cost of the chip," says Janak Patel of the University of Illinois at Urbana-Champaign. Patel designed some of the first testing and verification software. Today chips are being designed while keeping the requirements of testing software in mind. With the growth in complexity of chips, there is a corresponding growth in testing and verification software.

THE OTHER WONDERS

While the main application of semiconductors has been in integrated circuits, the story will not be complete without mentioning a few other wonders of the sand castle.

While CMOS has led to micro-miniaturisation and lower and lower power applications, the Integrated Gate Bipolar Transistors, or IGBT— co-invented by Jayant Baliga at General Electric in the 1970s—rule the roost in most control devices. These transistors are in our household mixers and blenders, in Japanese bullet trains, and in the heart defibrillators used to revive patients who have suffered heart attacks, to name a few applications. The IGBTs can handle megawatts of power. "It may not be as big as the IC industry but the IGBT business has spawned a billion-dollar industry and filled a need. That is very satisfying," says Jayant Baliga, who is trying to find new applications for his technology at Silicon Semiconductor Corporation, the company he founded at Research Triangle Park in Raleigh, North Carolina.

As we saw earlier, certain properties of silicon, such as its oxide layer, and the amount of research done on silicon have created an unassailable position for this material. However, new materials (called compound semiconductors or alloys) have come up strongly to fill the gaps in silicon's capabilities.

Gallium arsenide, gallium nitride, silicon carbide, silicon-germanium and several multi-component alloys containing various permutations and combinations of gallium, aluminium, arsenic, indium and phosphorus have made a strong foray into niche areas. "Compound semiconductors have opened the door to all sorts of optical devices, including solar cells, light emitting diodes, semiconductor lasers and tiny quantum well lasers," says Sorab Ghandhi, who did pioneering work in gallium arsenide in the 1960s and '70s.

"Tomorrow's lighting might come from semiconductors like gallium nitride," says Umesh Mishra of the University of California at Santa Barbara. He and his colleagues have been doing some exciting work in this direction. "A normal incandescent bulb lasts about 1,000 hours and a tube light lasts 10,000 hours, but a gallium nitride light emitting diode display can last 100,000 hours while consuming very little power," says IIT Mumbai's Rakesh Lal, who wants to place his bet on gallium nitride for many new developments.

Clearly, semiconductors have broken barriers of all sorts. With their low price, micro size and low power consumption, they have proved to be wonder materials. An amazing journey this, after being dubbed "dirty" in the thirties.

To sum up the achievement of chip technology, if a modern-day cell phone were to be made of vacuum tubes instead of ICs, it would be as tall as the Qutub Minar, and would need a small power plant to run it!

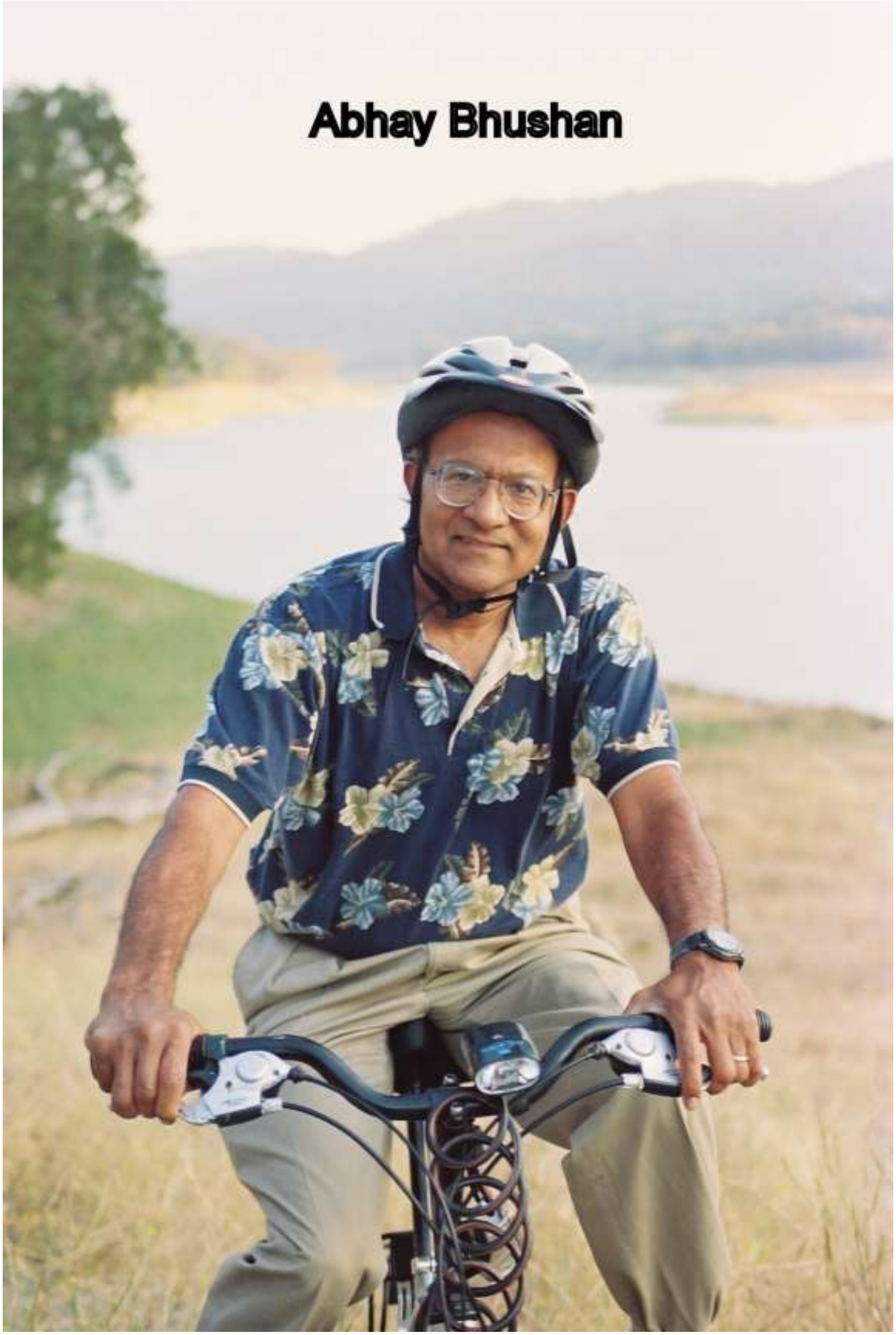
FURTHER READING

1.Nobel Lecture—John Bardeen, 1956 (http://www.nobel.se/physics/laureates/1956/bardee-lecture.html)
2. Nobel Lecture—William Shockley, 1956 (http://www.nobel.se/physics/laureates/1956/shockley-io.html)

3. The Solid State Century, *Scientific American,* Special issue, Jan. 22, 1998 Cramming more components onto integrated circuits—Gordon E Moore, *Electronics,* Vol 38, Number 8, April 19, 1965
4. The Accidental Entrepreneur—Gordon E Moore, *Engineering & Science*, Summer 1994, vol. LVII, no. 4, California Institute of Technology.
5. Nobel Lecture 2000—Jack Kilby (http://www.nobel.se/physics/ laureates/2000/kilby-lecture.html)
6. When the chips are up: Jack Kilby, inventor of the IC, gets his due with the Physics Nobel Prize 2000, after 42 years—Shivanand Kanavi, *Business India,* Nov. 13-16, 2000 (http://reflections-shivanand.blogspot.com/2007/08/jack-kilby-tribute.html)
7. From Sand to Silicon: Manufacturing an Integrated Circuit—Craig R. Barrett, *Scientific American,* Jan 22, 1998
8. The work of Jagdish Chandra Bose: 100 years of mm-Wave Research—D.T. Emerson, National Radio Astronomy Observatory, Tucson, Arizona (http://www.qsl.net/vu2msy/JCBOSE.htm)

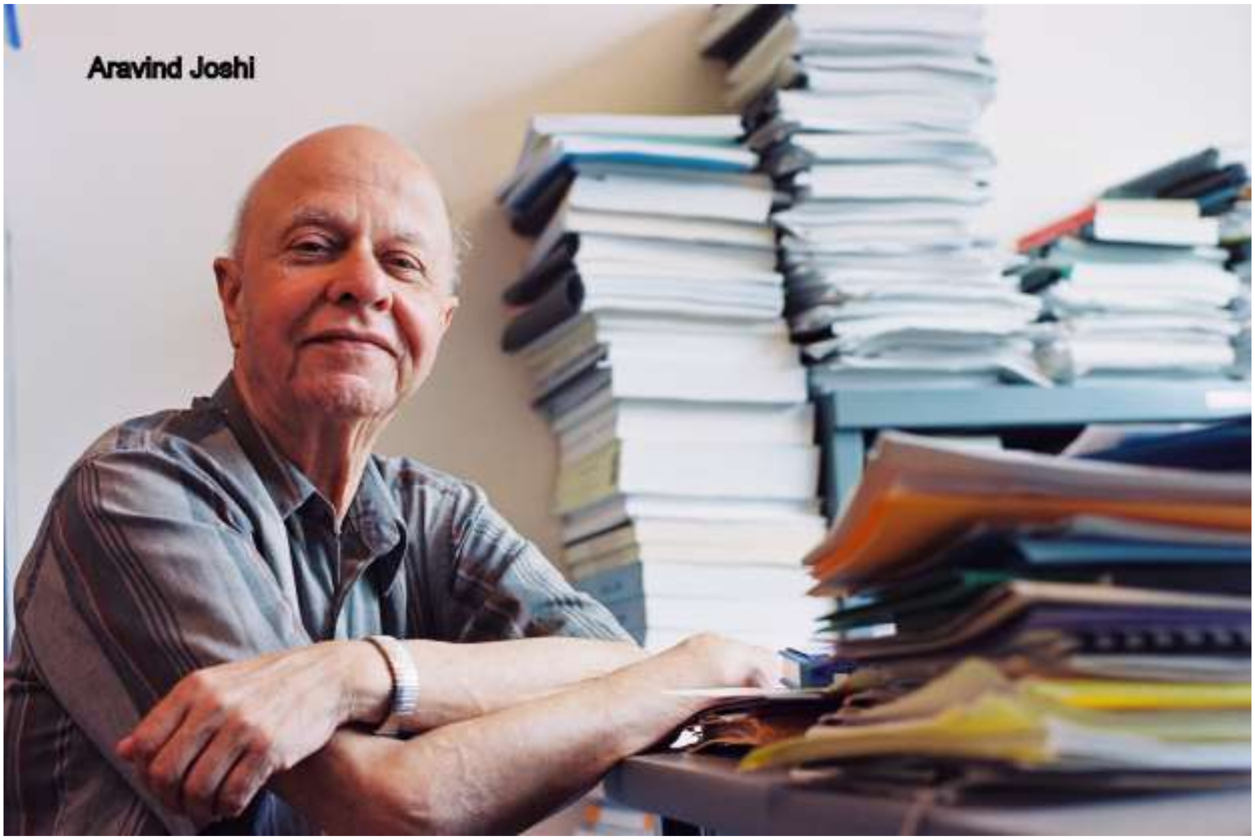
9. The Softening of Hardware—Frank Vahid, *Computer,* April 2003, Published by IEEE Computer Society

Abhay Bhushan

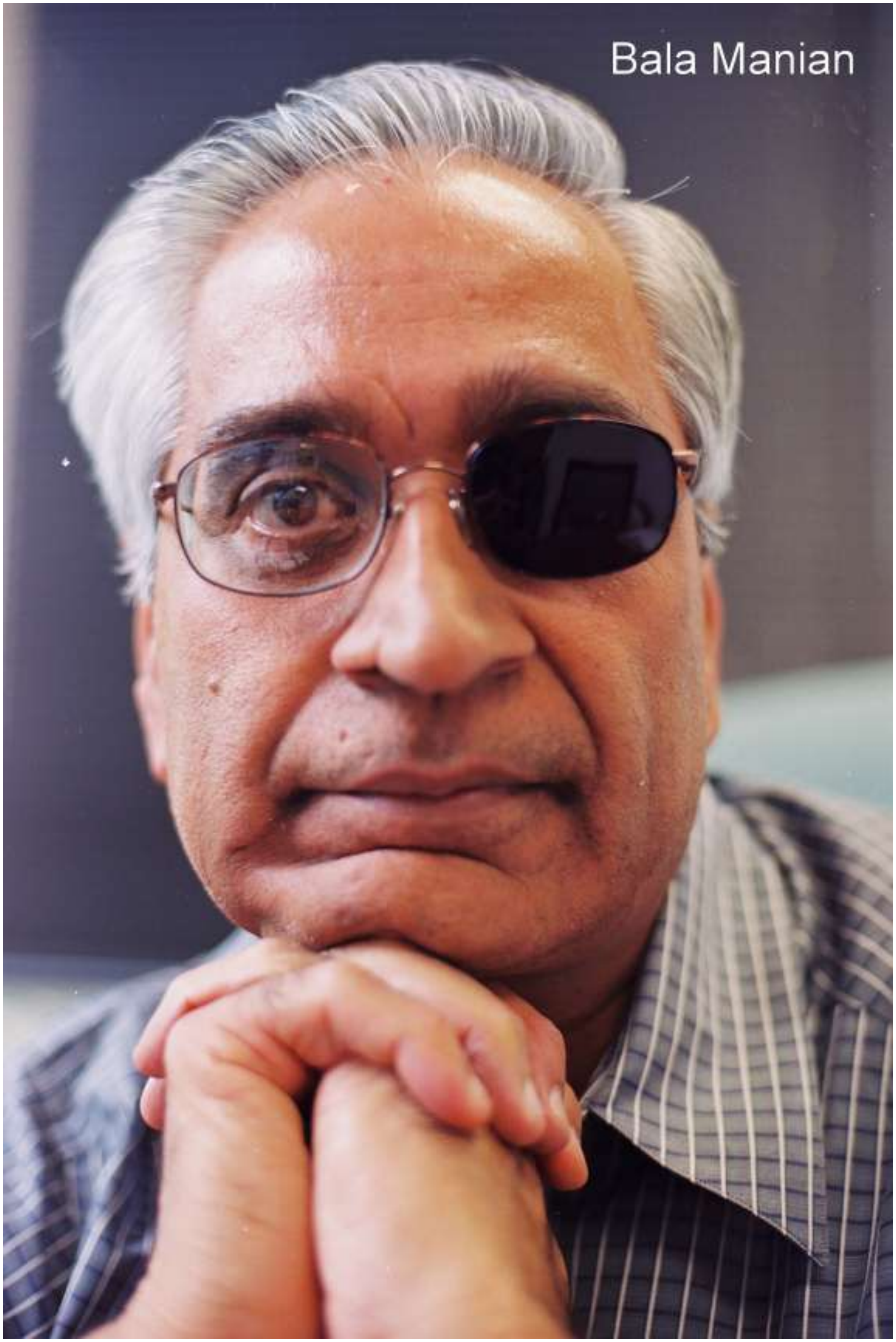



Amar Bose

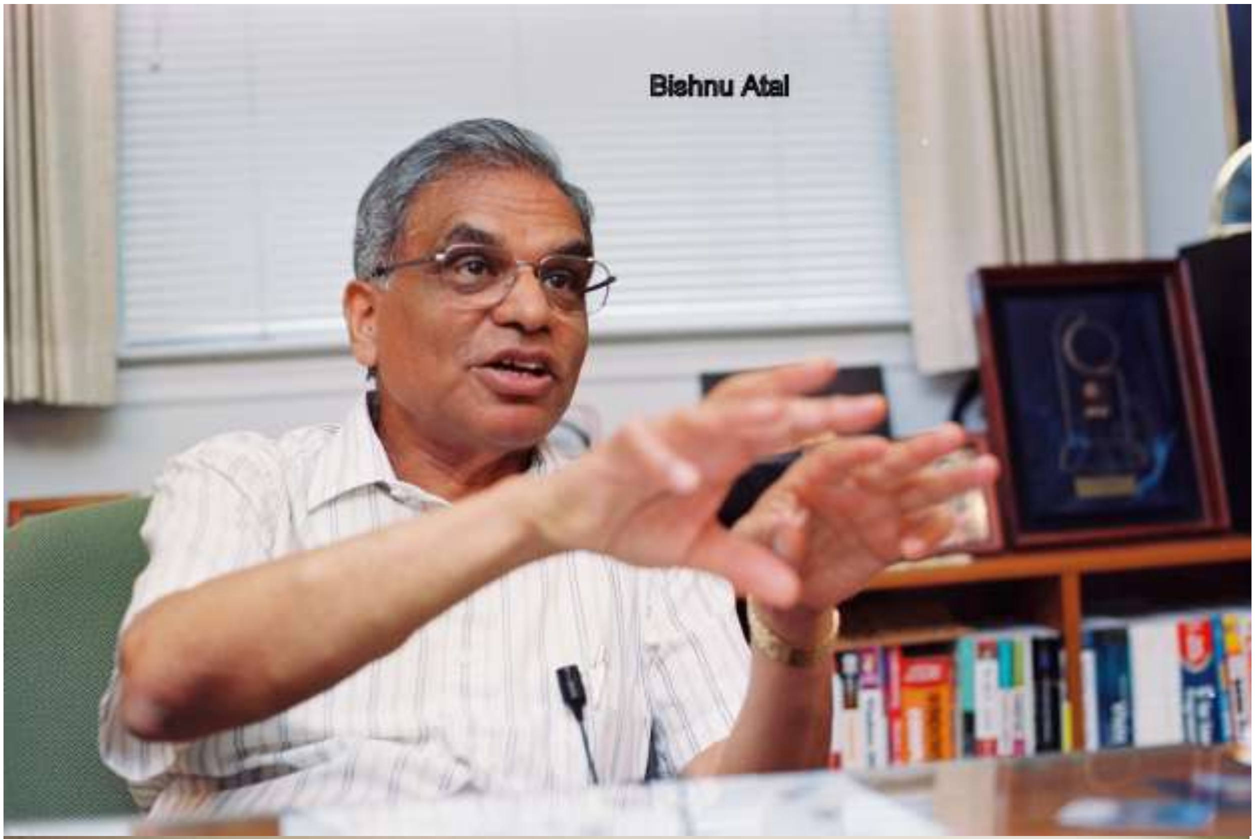

Arun Netravali



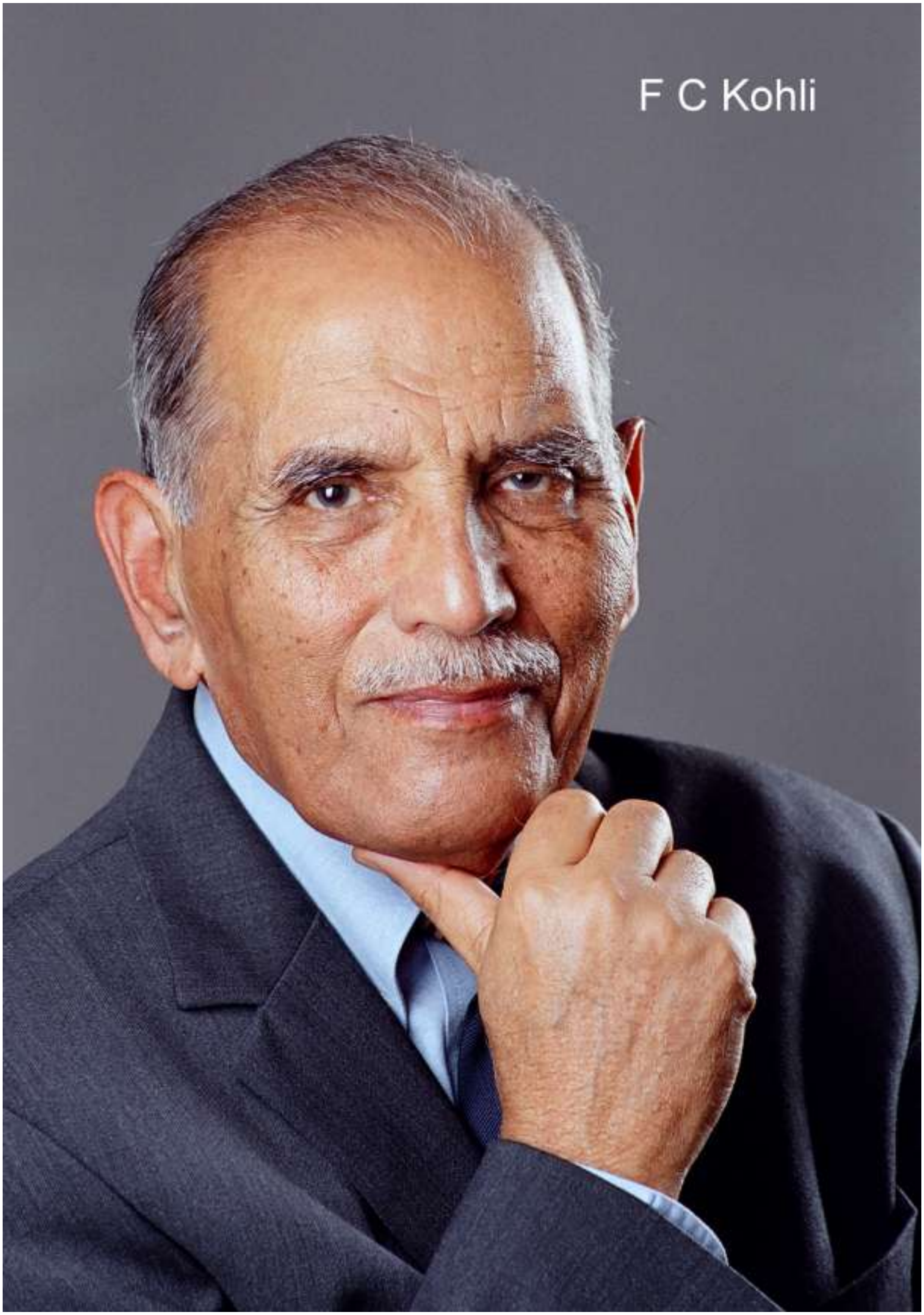
Aravind Joshi



Bala Manian

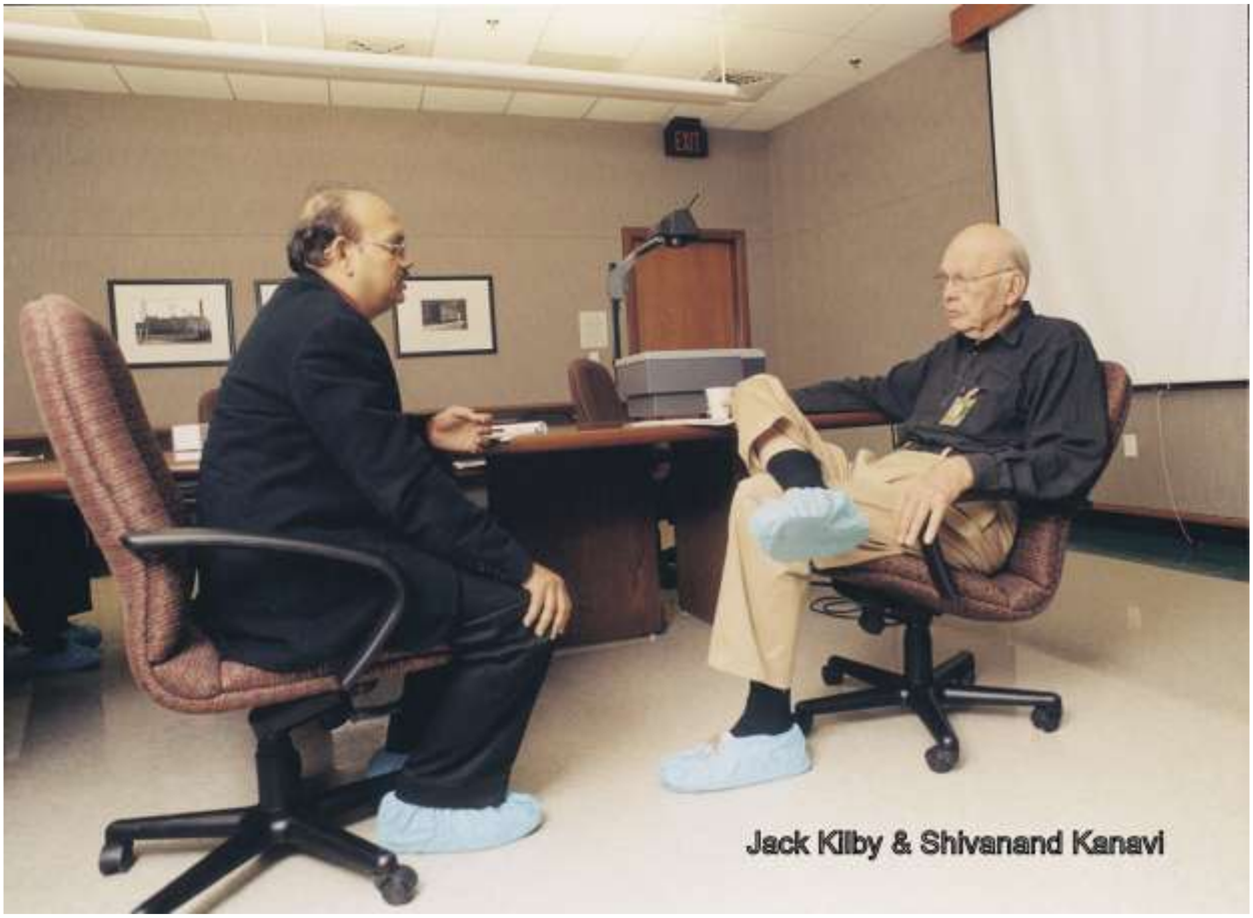

Bishnu Atal


Debasis Mitra

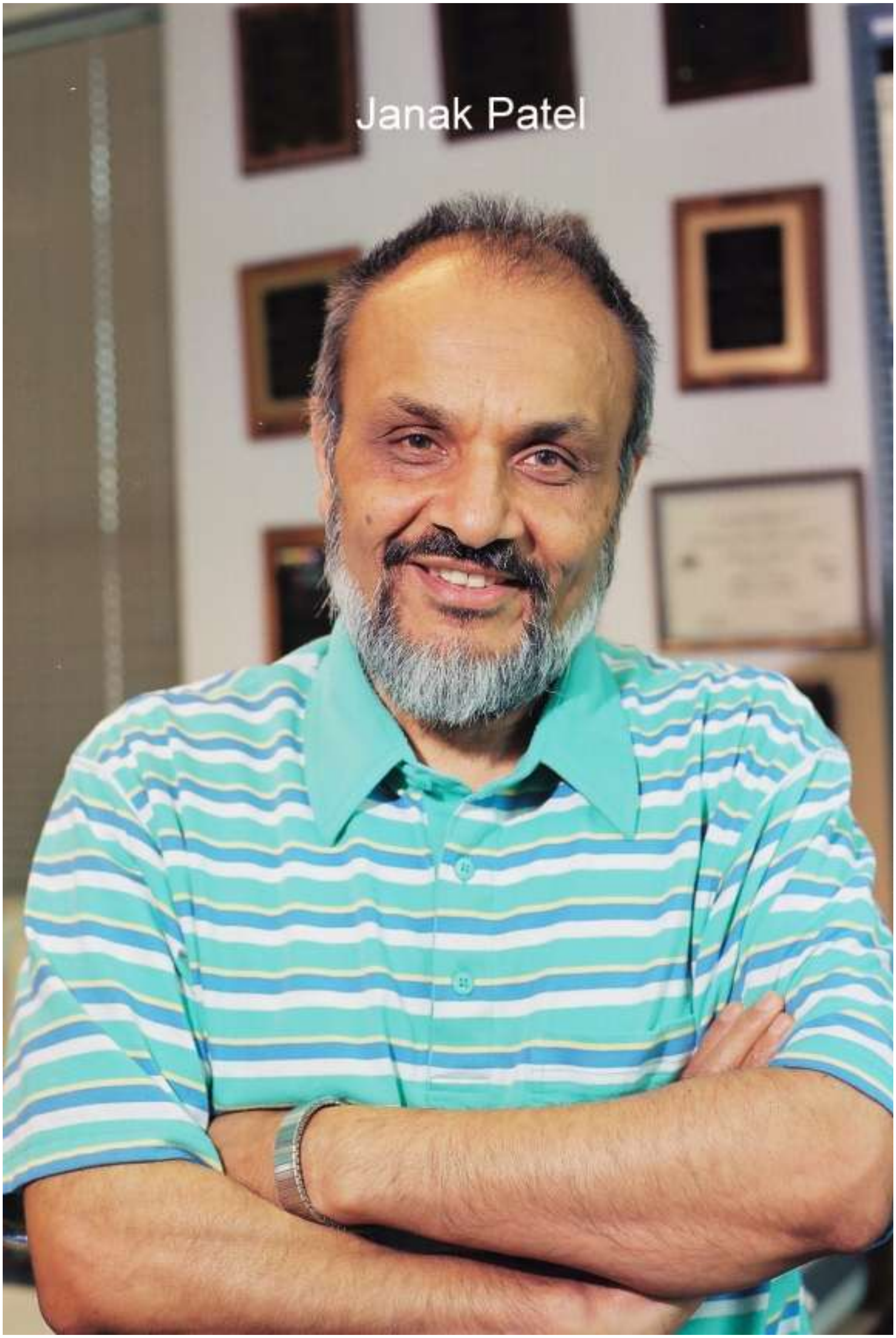


F C Kohli

Gururaj Desh Deshpande



Jack Kilby & Shivanand Kanavi



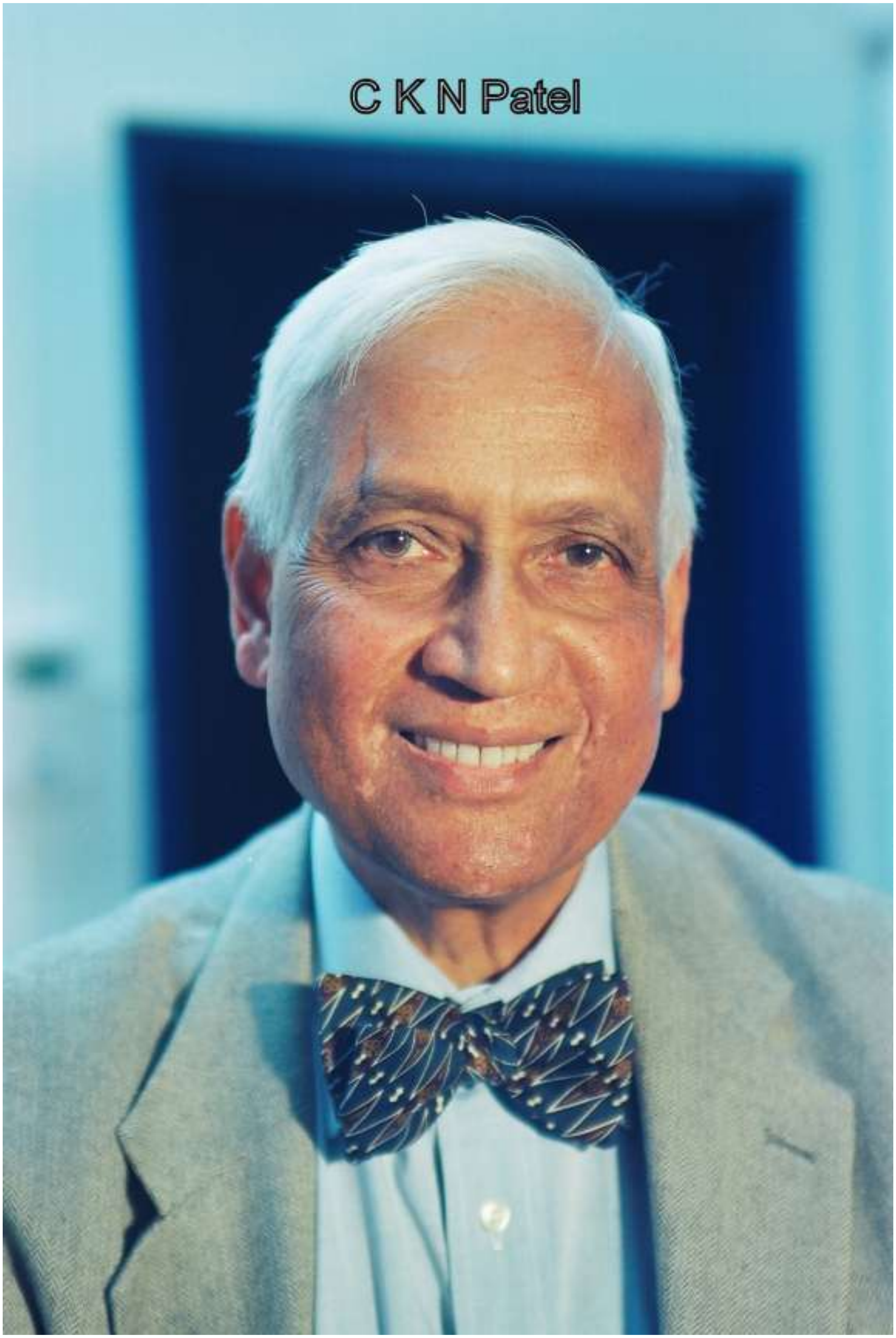
Janak Patel

Jay Menon

Jayant Baliga

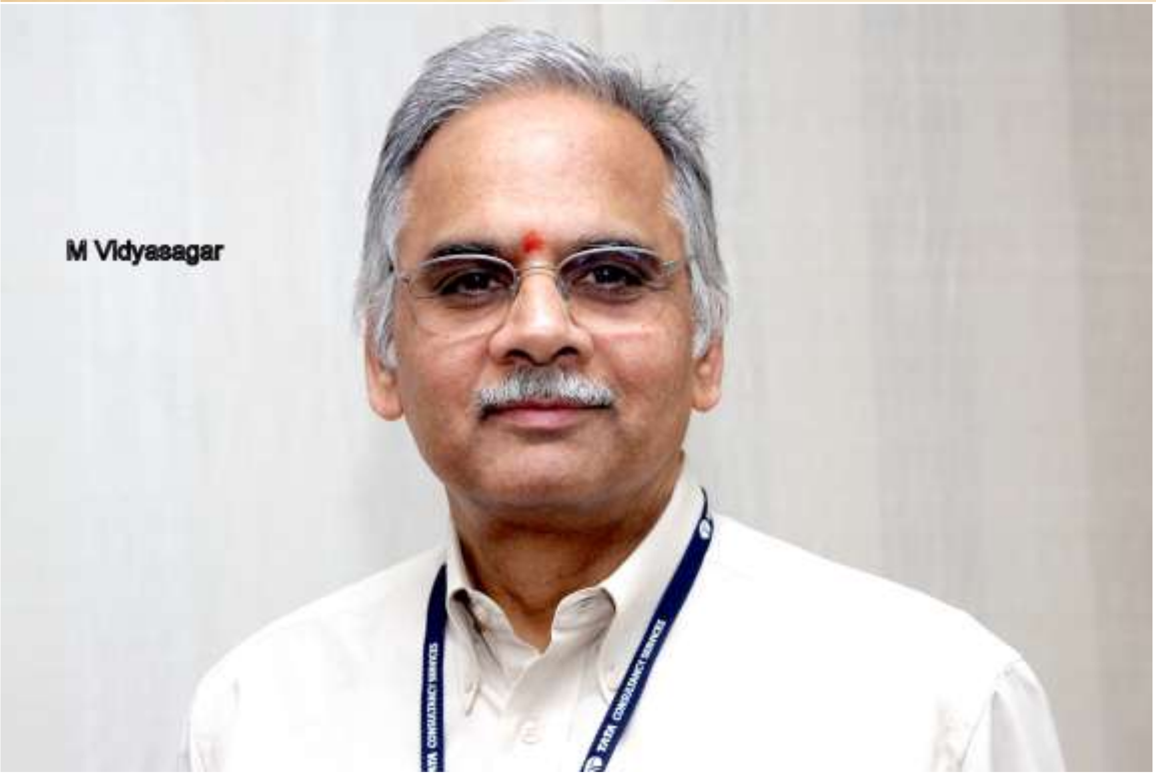
Jnan Dash

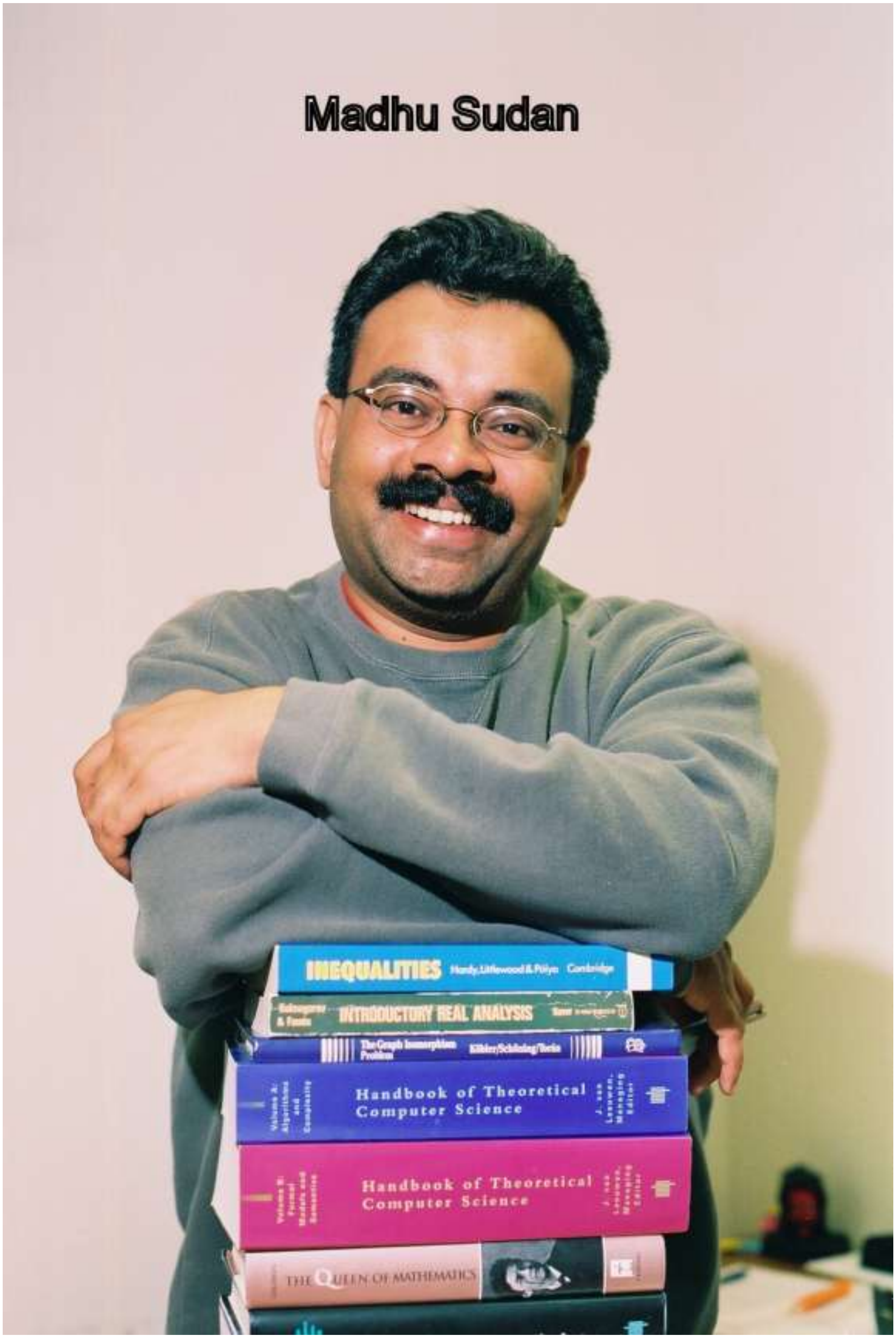

Kesav Nori



C K N Patel

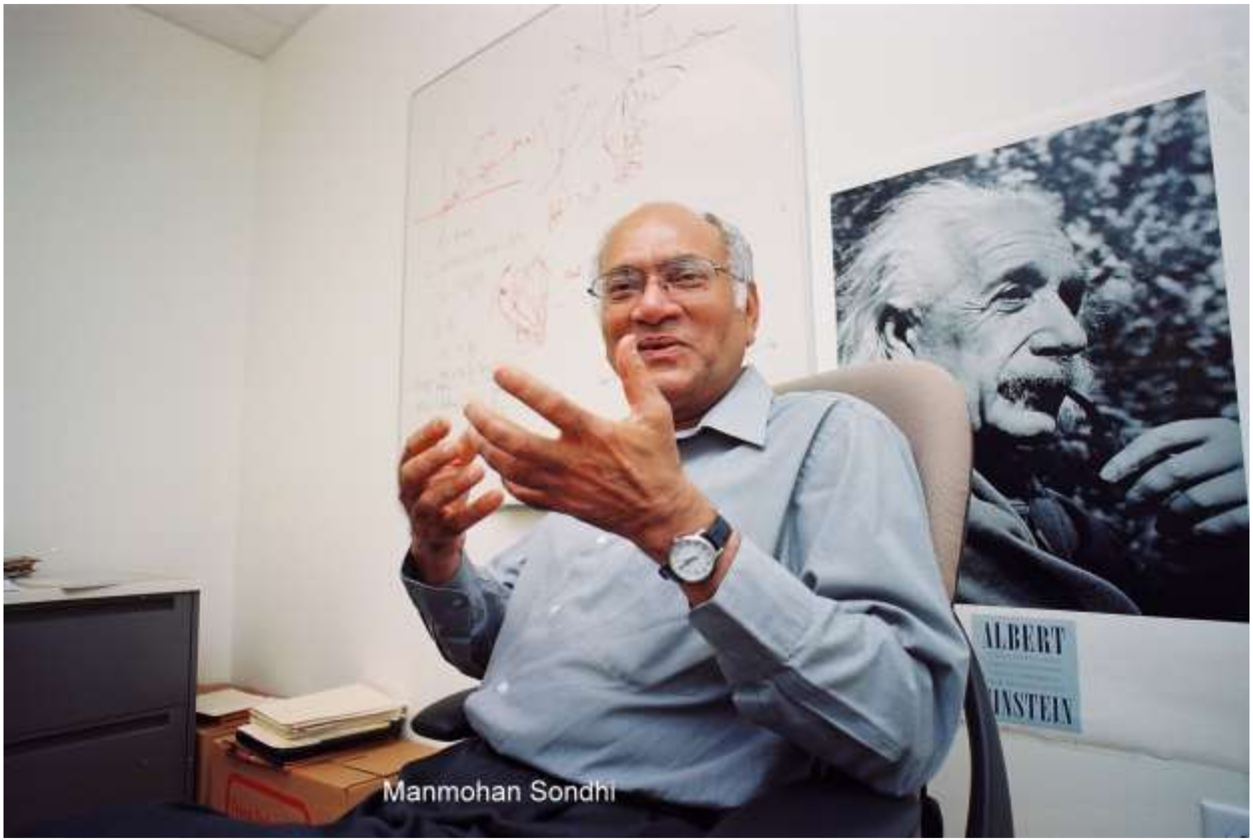

Luv Grover


M Vidyasagar



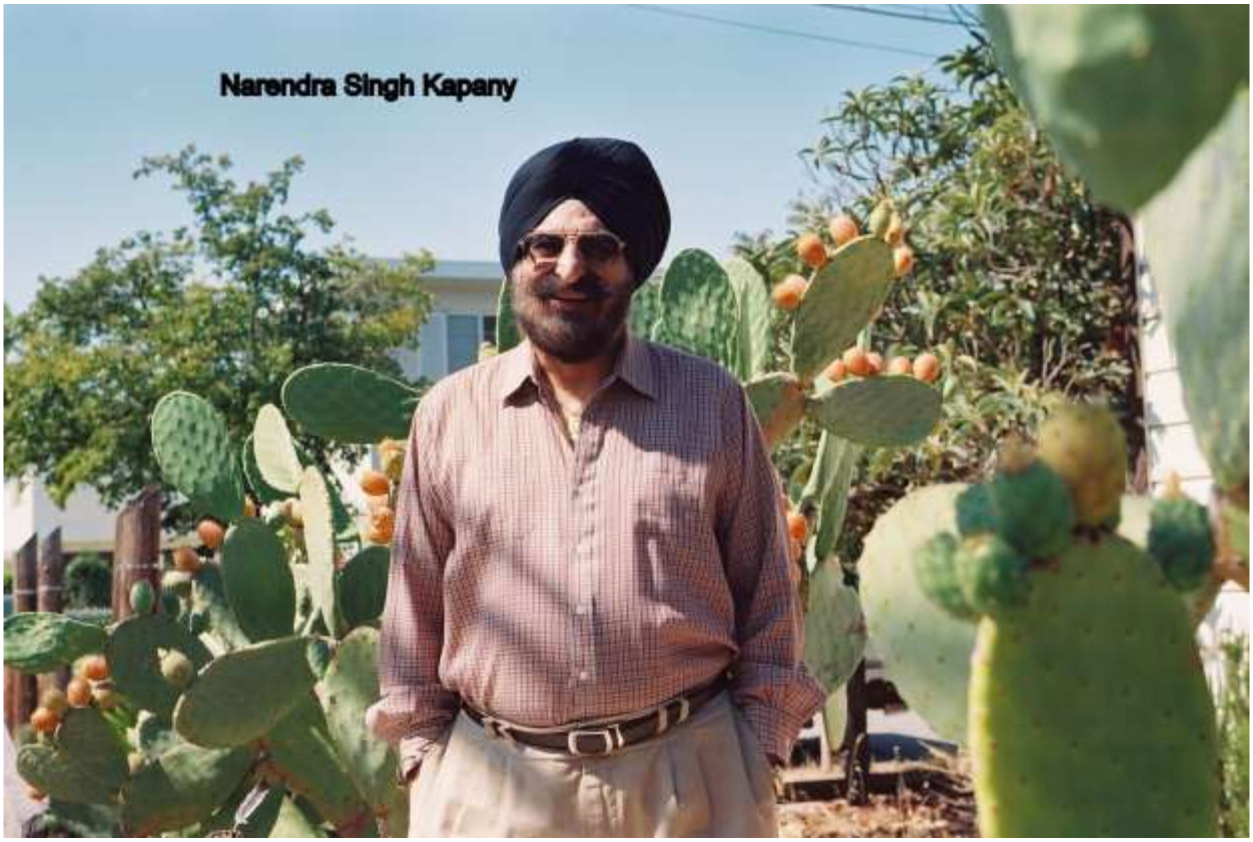
Madhu Sudan

INEQUALITIES Hardy, Littlewood & Pólya Cambridge
INTRODUCTORY REAL ANALYSIS
The Graph Isomorphism Problem Köbler/Schöning/Torán
Handbook of Theoretical Computer Science
Handbook of Theoretical Computer Science
THE QUEEN OF MATHEMATICS



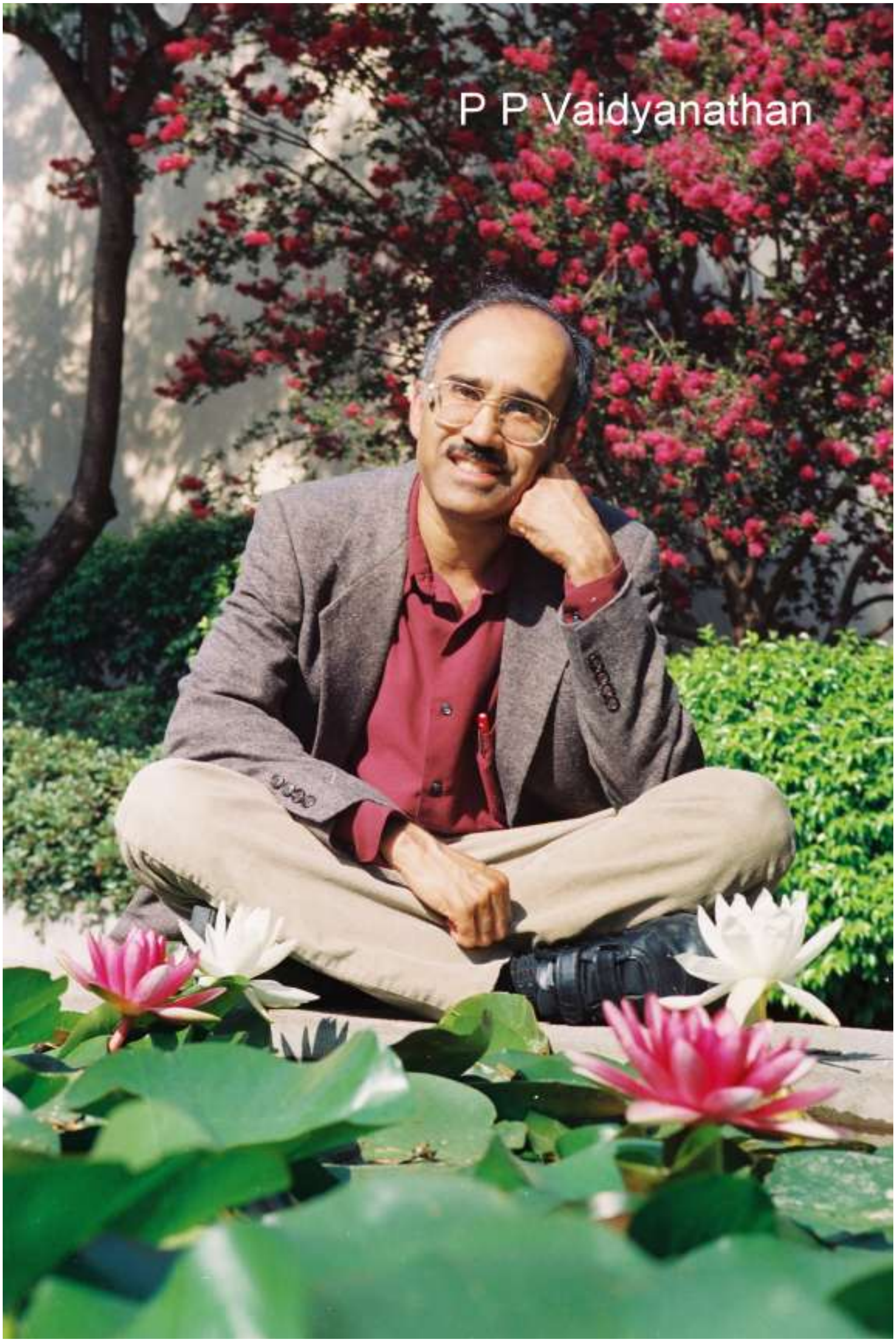
Manmohan Sondhi

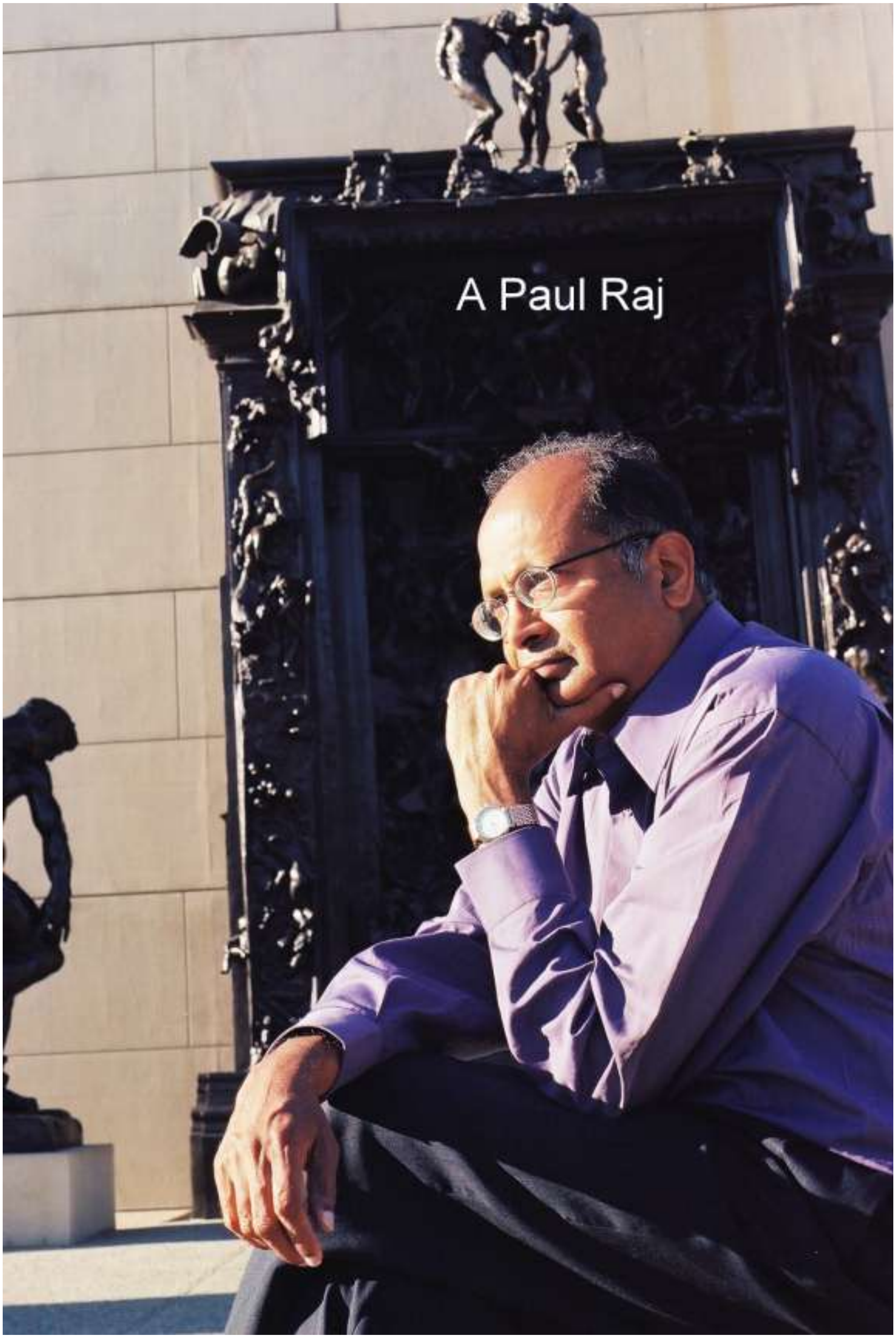
Mriganka Sur


Narendra Singh Kapany

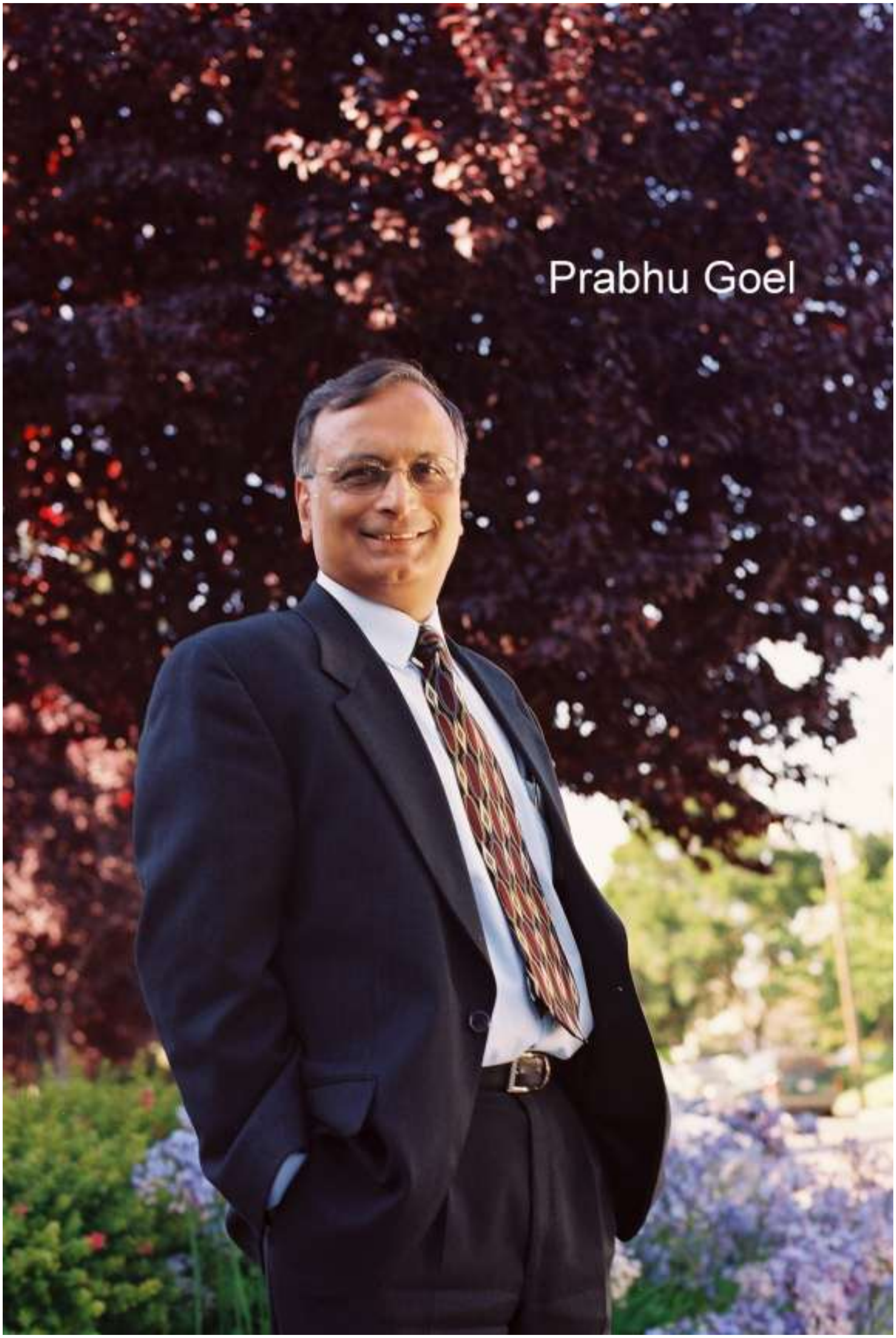

Narendra Karmarkar

Nikil Jayant


P P Vaidyanathan



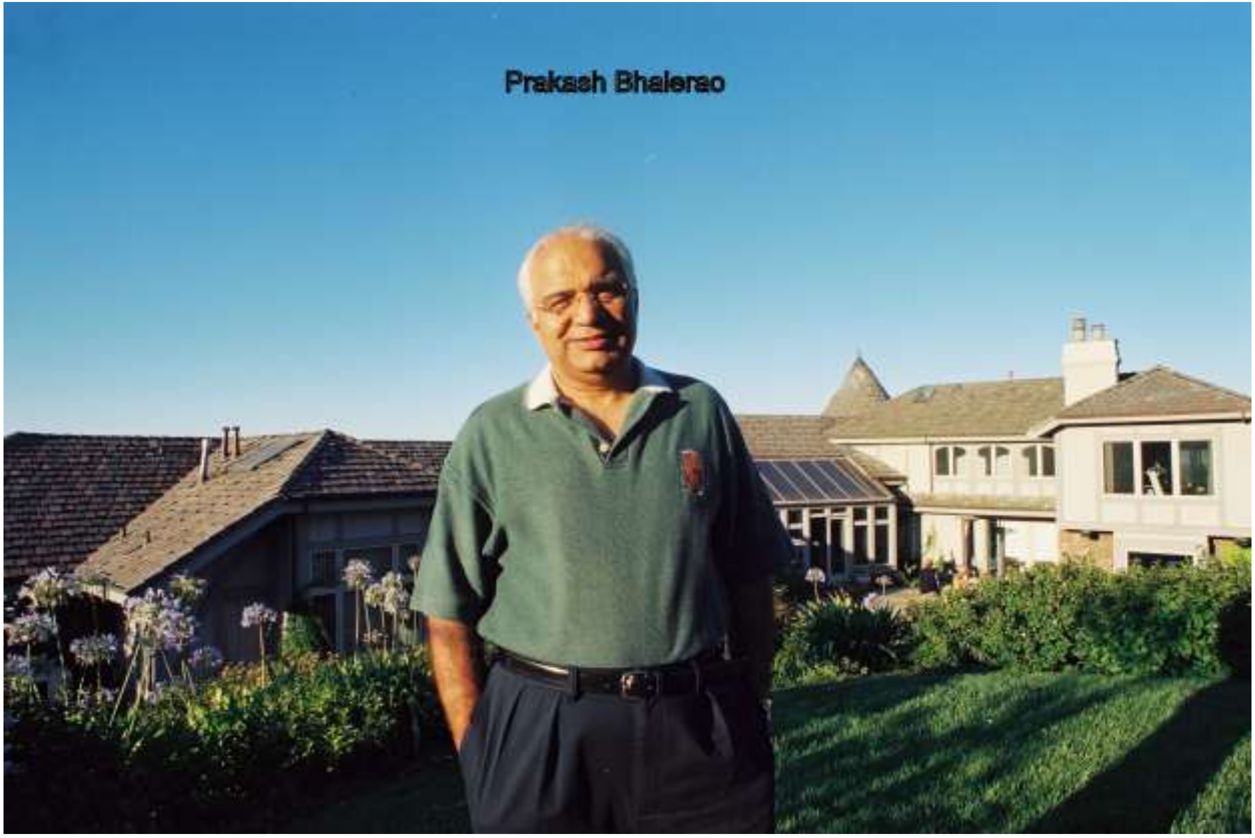
A Paul Raj


Prabhu Goel


Pradeep Khosla

Pradeep Sindhu

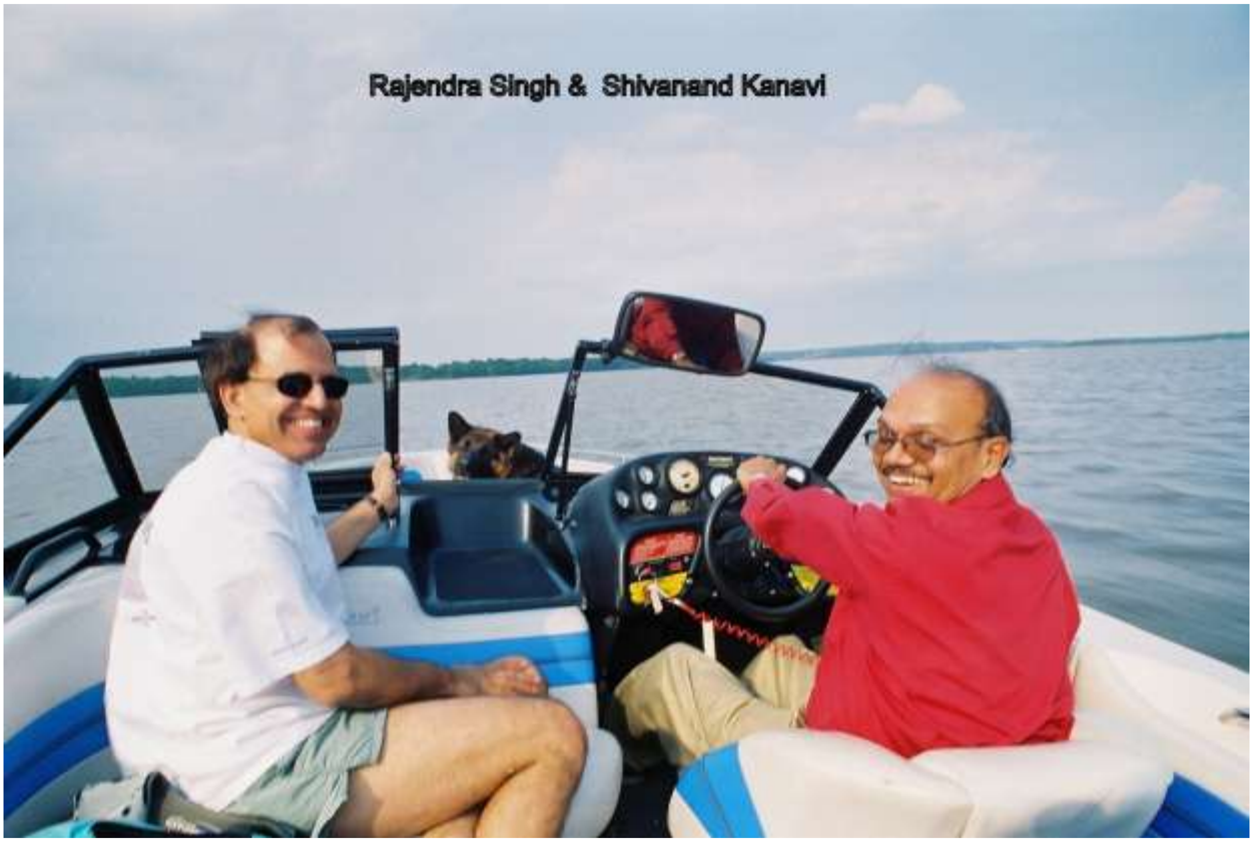

Pradman Kaul


Prakash Bhalerao

Praveen Chaudhari


Raj Reddy

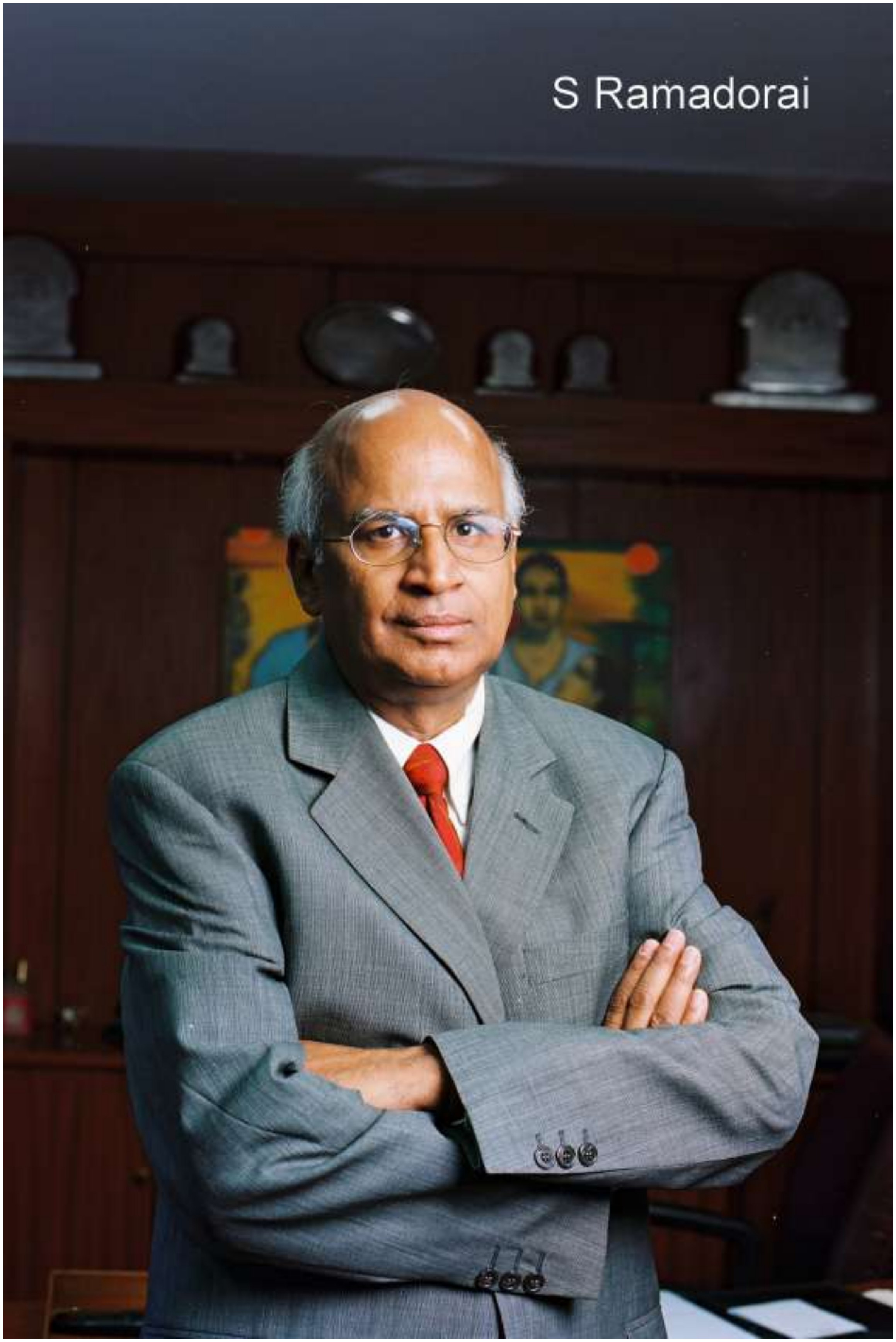
Rajendra & Neera Singh



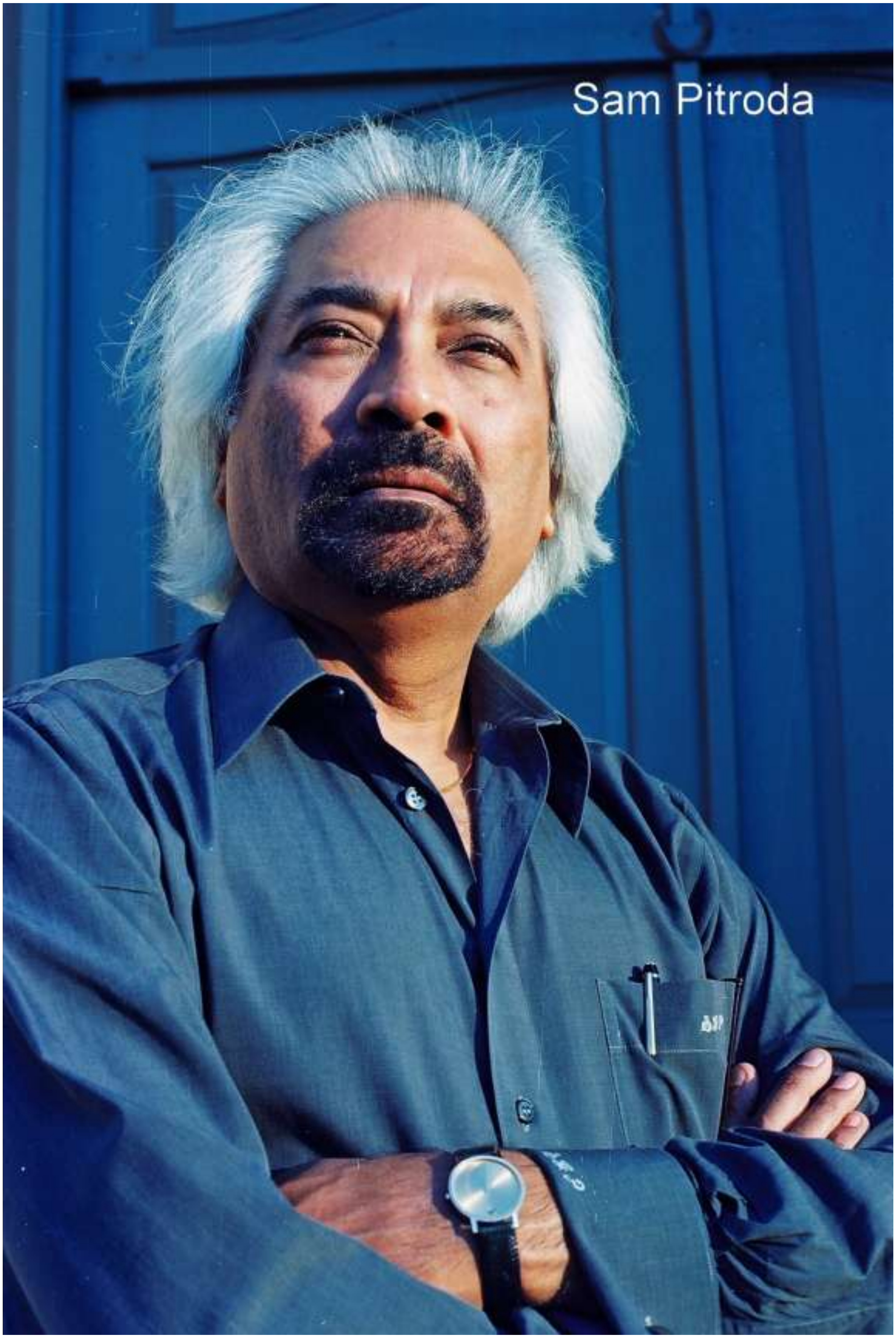
Rajendra Singh & Shivanand Kanavi

Rajeev Motwani

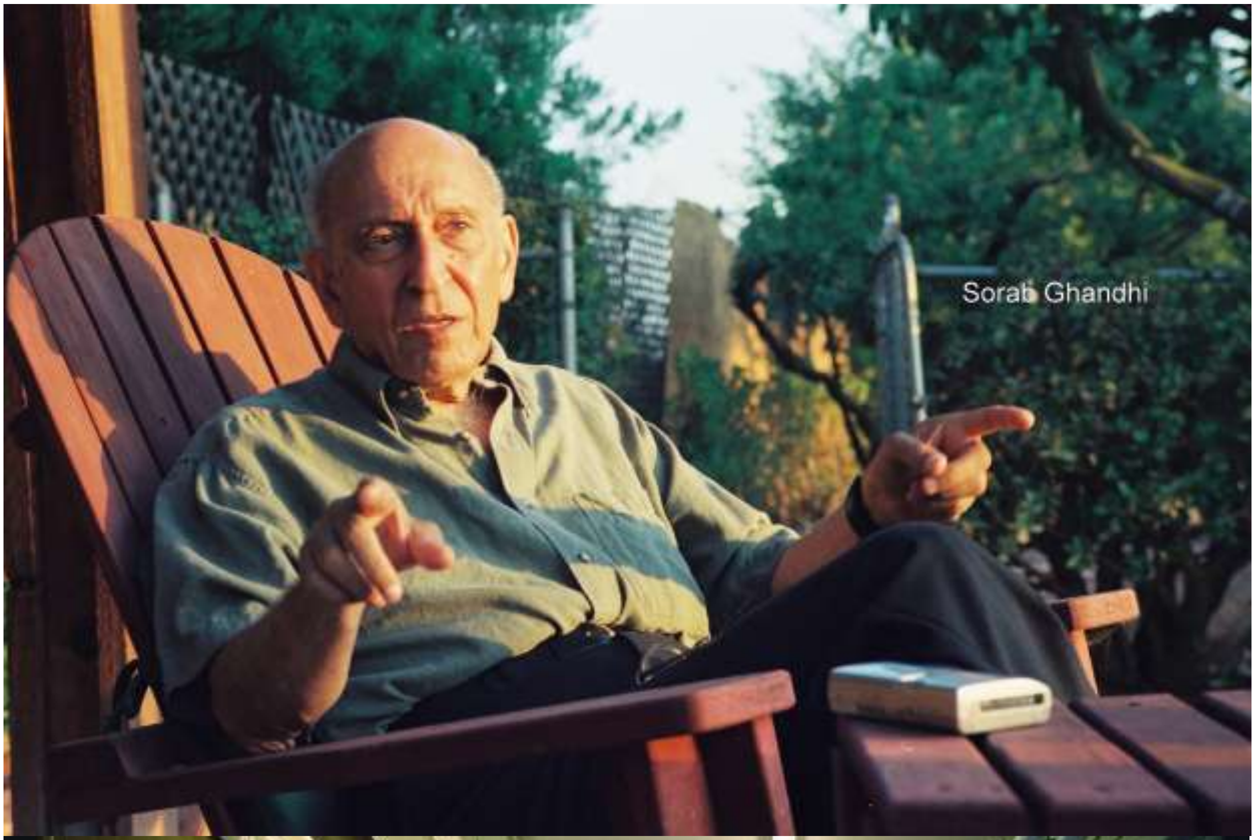

Ravi Kannan



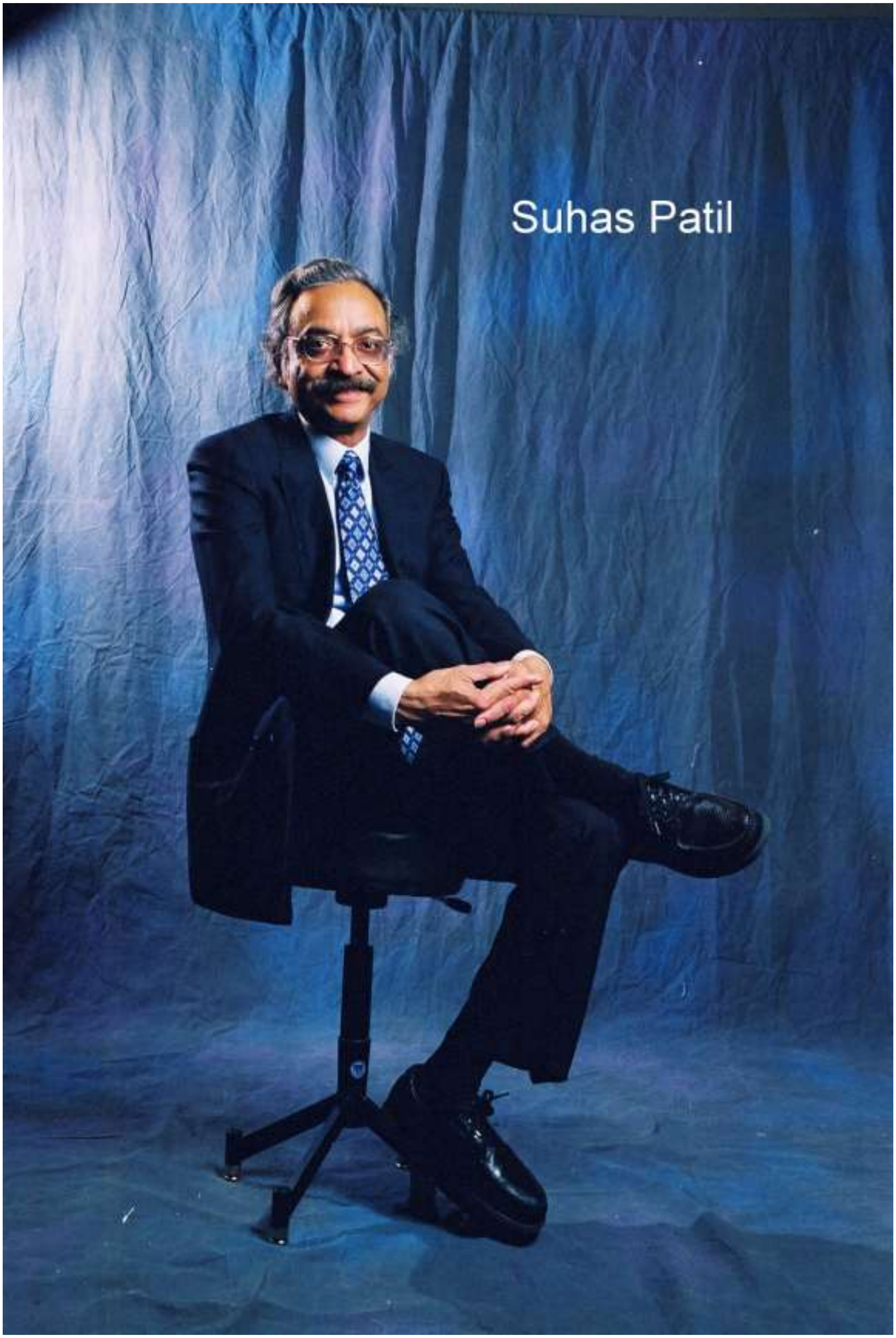
S Ramadorai


Sam Pitroda


Sanjit Mitra

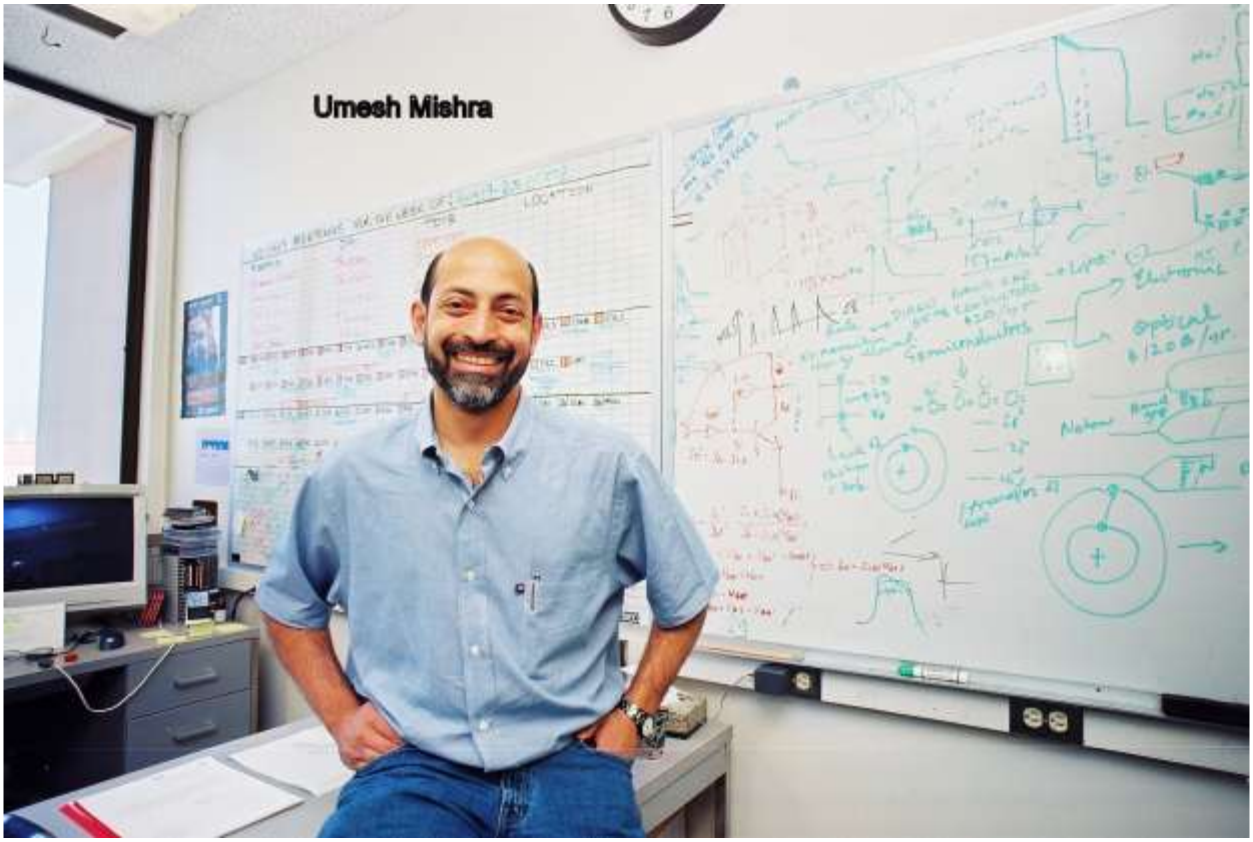

Sorab Ghandhi


Subra Suresh



Suhas Patil

Tom Kailath

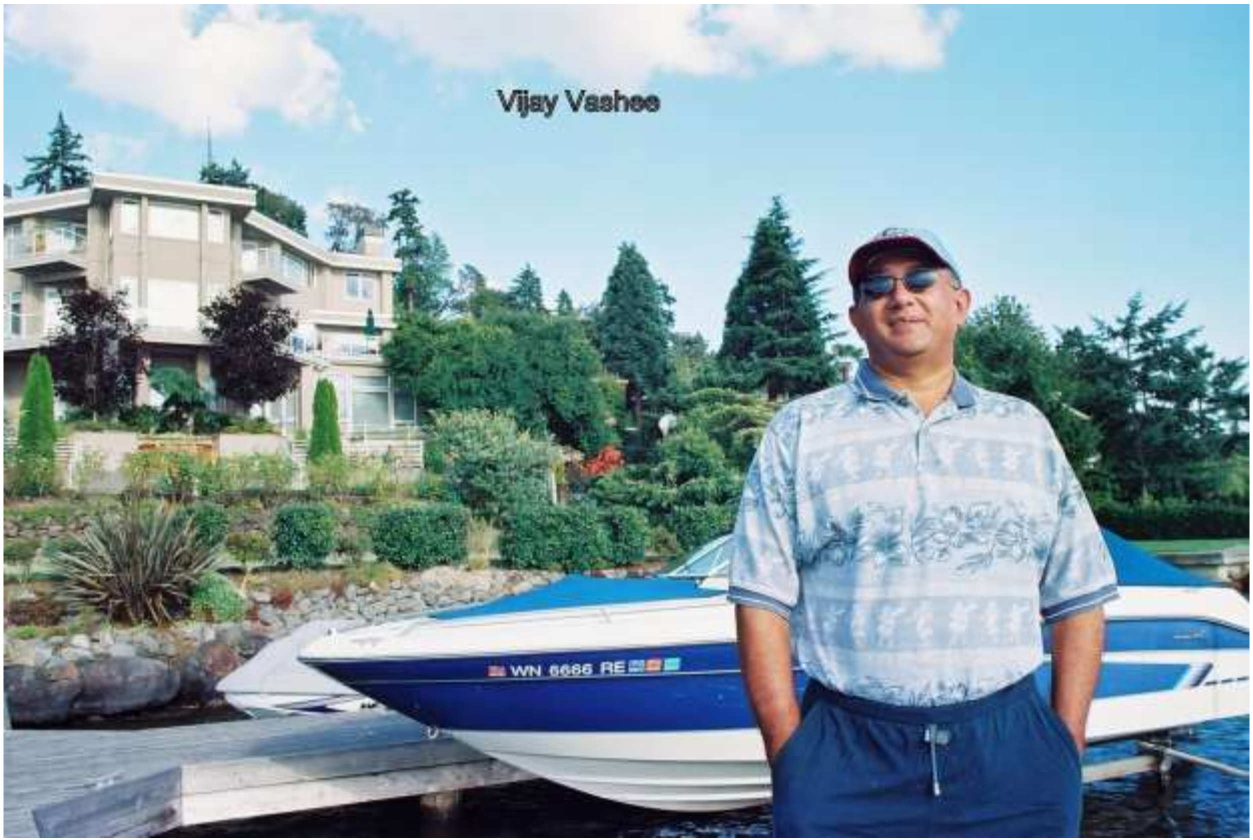


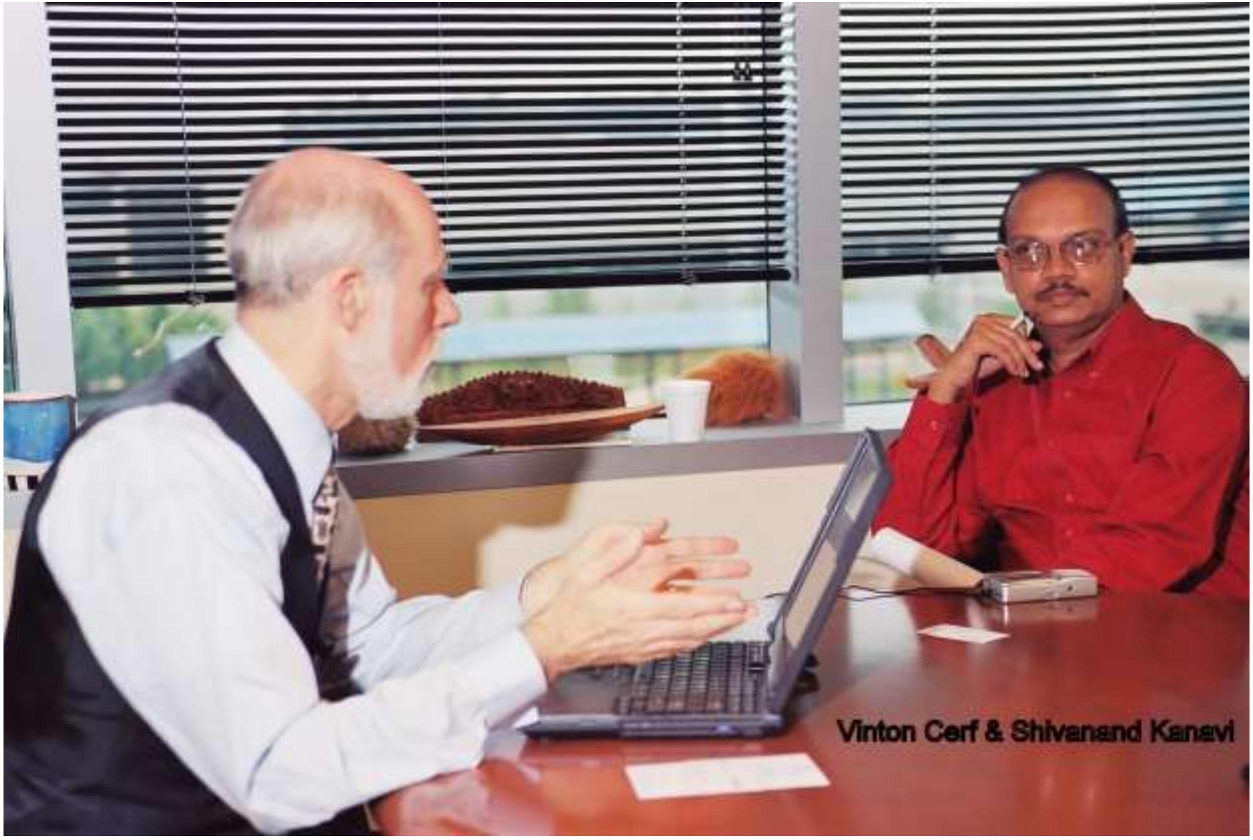
Umesh Mishra

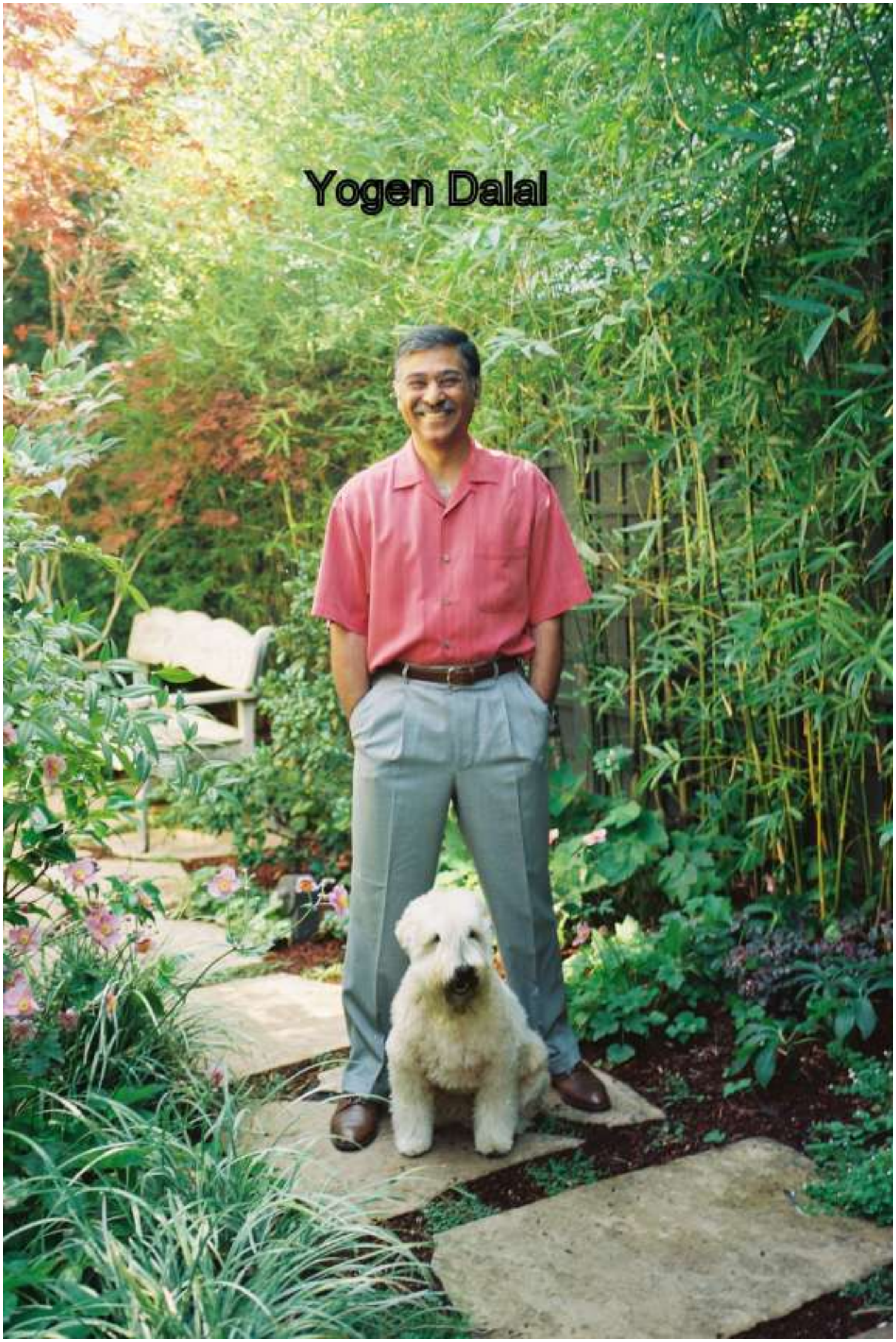

Umesh Vazirani

V Rajaraman


Venky Narayanmurthy


Vijay Vashee


Vinton Cerf & Shivanand Kanavi


Yogen Dalal

COMPUTERS: AUGMENTING THE BRAIN

"One evening I was sitting in the rooms of the Analytical Society at Cambridge, in a kind of dreamy mood, with a table of Logarithms open before me. Another member, coming into the room and seeing me half asleep, called out 'What are you dreaming about, Babbage?' I said, 'I am thinking that all these tables may be calculated by machine."

—**CHARLES BABBAGE**, (1792-1871)

"The inside of a computer is as dumb as hell but it goes like mad! It can perform very many millions of simple operations a second and is just like a very fast dumb file clerk. It is only because it is able to do things so fast that we do not notice that it is doing things very stupidly."

—**RICHARD FEYNMAN,** PHYSICS NOBEL LAUREATE, (1918-1991).

Computers form the brain of the IT revolution. They are new genies that are becoming omnipresent in our life; helping us make complex decisions in a split second, be it in a factory or at a space launch; controlling operations of other highly complex machines like a huge airplane, or a machine tool; carrying out imaginary scientific experiments by simulating them; making our fantasies come true in the movies with mind-boggling special effects; they even help me write this book and edit it, and so on.

According to a cover feature in the February 2003 issue of IEEE Spectrum magazine, a modern car has on the average fifty computers inside, which control fuel injection, ignition, traction, braking, air-bag, diagnostics, navigation, climate control and in-car entertainment. Computers are rapidly growing in the capacity to store information and calculate numbers with increasing speed. The magical realm of computers is extending its frontiers by the day.

All this is making many laymen believe in the computers' omnipotence and omniscience. As usual film scriptwriters are having a field day with apocalyptic visions of manmade machines taking over control of human kind. A glimpse of that was provided in the Arthur C Clark-Stanley Kubrik sci-fi classic of the sixties—*2001 Space Odyssey*—where a computer—HAL—takes over control of the space mission, to the more recent *Terminator*. All of them portray computers as intelligent Frankensteins.

"Oh, that is good for an evening with popcorn, but computers are just tools," says the sophisticate. "But computers are not just tools of the old kind," says Kesav Nori of Tata Consultancy Services, whose work on Pascal compiler is well known. "Man has been making tools and harnessing energy since the agricultural revolution. Then came the Industrial Revolution, and now it is legitimate to talk of a new revolution—the information revolution," he says.

The IC and the microprocessor, which we reviewed in the first chapter, have fuelled the affordability and the power of computers. However, computers owe their theoretical origins to a convergence of three streams of thought in logic, mathematics and switching circuits, spanning over three centuries. We will run through at a trot to absorb the main ideas.

THE INFORMATION REVOLUTION

What is information revolution? Our capacity to store, communicate and transform information has been growing since millennia. We had flat stones and paper to store information earlier and now we have magnetic and optical devices: tapes, floppies, hard disks, CDs, DVDs and so on.

We also developed a device to communicate—language—and a script to record it. Storing information or our thoughts and reflections on nature (including ourselves), made it easier to transport the content to another place or another time. Printing press was a big step forward.

Transporting information has evolved from physical carriers like monks and traders or messengers to non-human carriers like pigeons, heliograph, telegraph, telephone, radio, television and now the Internet.

In order to understand information, to convert it into knowledge and wisdom we have been using our brain and continue to do so. Now new devices like calculators and computers have been invented to help our brains perform some information processing tasks.

However, there is one qualitative difference between computers and other tools. Computers are universal. What does it mean?

VISHWAROOP† OF THE COMPUTER

Each machine of the great industrial revolution like a loom, a motor, a lathe and so on take in energy and certain raw materials as input, transform it in a specified way and give us outputs. Thus we know the outputs of a loom or a lathe or a motor. The machines can be refined, made more efficient, reliable over repeated operations and so on. Such machines powered manufacturing and were the hallmark of the great Industrial Revolution. But they are all special purpose machines.

The computer invented by Alan Turing and John von Neumann in the 1930s and '40s, is radically different. It can simulate any machine. It is a general-purpose machine. How is that possible? How can one machine act like a loom, a lathe, an airplane and so on? The reason is that *most processes can be modelled*: be they editing a text; drawing a picture; weaving a pattern; piloting an airplane; converting a machine drawing to metal cutting; doing 'what if' analysis on budgets and sales-data and so on.

Once modelled, we can write logical programmes simulating them. These logical programmes are converted into binary codes of '0' and '1' and processed by the computer's circuits. After processing, the computer would give the result of what the lathe would have actually done, given a certain input. This *mimicking is called simulation*.

---

†A divine revelation that contains the entire universe, Universality.

Once we are convinced that it is simulating the lathe repeatedly and accurately, then we can use it as a mechanical brain to actually instruct the lathe, what should be done next to achieve the desired result. It becomes a 'controller'.

The same computer can be programmed to simulate a loom and then run a loom instead of a lathe and so on. Computer scientists call this property, 'universality'. When this *vishwaroop* of the computer hits you then one realises that there is something revolutionary that has happened and that is the basis of information revolution.

Of course, we still do not know how to logically model and simulate many phenomena and computers cannot help you there. Many of the activities of the human brain that we normally associate with 'consciousness'—self-awareness, emotions, creativity, dreams, cognition and so on—fall in this category.

COCKROACHES AND ARTIFICIAL INTELLIGENCE

Can machines get more and more powerful and surpass human beings in their abilities? This is a subject of deep research among engineers and extensive speculation among futurologists and pop philosophers. Feelings run high on this subject.

Computers have long ago surpassed human abilities in certain areas like the amount of information you can store and recall, or the speed with which one can do complex mathematics. Even in a game like chess, considered an intelligent one, they have beaten grand masters. But they have a very long way to go in any field that involves instinct, common sense, hunch, anticipation, creative solution etc. Bob Taylor, who was awarded the National Technology Medal by the US president in 2000 for his contributions to the development of Arpanet (precursor of the modern Internet) and Personal Computing, remarked to the author, "After fifty years of Artificial Intelligence (AI), we have yet to recreate the abilities of a cockroach, much less human intelligence!" Strong words, these.

For an assessment of the achievements and challenges before AI one can read Raj Reddy's Turing Award Lecture in 1995, *'To Dream The Possible Dream'.*

Be that as it may, we will look at computers in the rest of this chapter from a pragmatic viewpoint as new tools that can lighten the burden of our tedium and enhance our physical and mental capabilities—as augmenters of our brain and not replacements of it.

As we noted in the first chapter, the semiconductor microelectronics revolution has multiplied everything that was possible fifty years ago by a factor of a million, while simultaneously dropping the price of this performance. This phenomenon of *increasing performance and falling prices has led to exponential rise in the use of the new technology—Information Technology*. For discerning observers and social theorists any exponential characteristic in a phenomenon shows a revolution lurking nearby, waiting to be discovered.

COMPUTERS ARE DUMB MACHINES

Actually, computers are dumb machines. That is not an oxymoron. As Nobel Laureate Richard Feynman puts it in his inimitable style in *Feynman Lectures on Computation:* 'For today's computers to perform a complex task, we need a precise and complete description of how to do that task in terms of a sequence of simple basic procedures—the 'software'—and we need a machine to carry out these procedures in a specifiable order—this is the 'hardware'. In life, of course we never tell each other *exactly* what we want to say; we never need to. Context, body language, familiarity with the speaker and so on, enable us to 'fill in the gaps' and resolve any ambiguities in what is said. Computers, however, can't yet 'catch on'. *They need to be told in excruciating detail exactly what to do.'* The exact, *unambiguous recipe for solving a problem* is called an algorithm. It is a term of Arabic origin and is actually named after the famous Arab mathematician of the ninth century: Al Khwarizimi. This astronomer, mathematician from Baghdad introduced the Indian decimal system and algebra (another term derived from his book *Al Jabr*) to the Arab world. Interestingly his work was inspired by an astronomer–mathematician from India—Brahmagupta. When Khwarizimi's books were translated into Latin in the twelfth century they greatly influenced European mathematics.

Let us look at the process of multiplying ten by twelve. It is equivalent to ten added to itself twelve times, and we get the algorithm to multiply integers. While this is understandable, it is not intuitive that we can reduce non-numerical problems like editing a letter or drawing a picture or composing a musical piece to a set of simple mathematical procedures.

Computer scientists have been able to do that and that is leading to the continuous expansion of the realm of computers.

A computer user may not have the mathematical sophistication or the time and inclination to write an algorithm for a particular task like editing prose. So we break up the task into sub-tasks like deleting a word, cutting and pasting a piece of prose elsewhere or checking the spelling of what we have typed and so on. We can then leave the task of turning these commands into mathematical algorithms to the more sophisticated programmers. We thus create layers of programmers who convert a command into more and more involved mathematical and logical procedures.

The symbols used with a set of rules called the syntax or grammar to convert a task into algorithms form a programming language. There are other programs that convert the programmes written in programming languages into instruction to the machine to add two numbers, store the result somewhere, compare the result with a number already stored in some corner and so on. The computer's electronic circuits carry out these operations. They can add, store, and compare voltage signals representing '0' and '1' and give a result, which is again converted back by the programme into an understandable output like deletion of a word in this chapter.

THE ONLY GOOD COMPUTER IS A DUMB COMPUTER

"It is good that computers are dumb," says Kesav Nori. "Only then you can have repeatability and reliability. The programmes do what they are supposed to do, with no surprises. The models that we make in our brain are approximate. Moreover, the computer is a finite machine as opposed to the abstractions of the infinite in our brains. Realising this and developing efficient and reliable ways of simulating the model through an algorithm is what programming is all about," says he.

The innards of a computer consist of a way to give inputs to the computer, a place to store information called the memory, a place to do various operations like add, store and compare information called the processor and a way to express the results called the output. This forms the hardware of all computers, be they giant supercomputers or a small video game machine.

The drive to make these innards smaller, faster, less power hungry and cheaper has led to the evolution of the hardware industry. The urge to split various numerical and non-numerical tasks into simple mathematical operations that can be carried out by the hardware, has led to the software industry. The hardware and software industries, working in tandem, are creating affordable computers that can do increasingly sophisticated tasks.

INSCAPE: COMPUTERS AS FILE CLERKS

Inside the computer, we see a really busy machine. The computer computes for only a small part of the time, while most of the time it is storing data, retrieving data, copying data to another location and so on. Thus if data can be compared to office files, then it looks like a whole bunch of clerks busy shuffling paper inside the computer.

Let us say we are inside a big company where all kinds of sales data have come into the head office from different sources and there is a filing clerk who knows how to store and retrieve a file.

He has written each piece of sales figure on a card, recorded which salesman did the sale, the location of the sale, etc. Now, a bright young executive, who has to submit one of those endless reports to senior managers, asks the clerk, "what is our total sales in Mumbai, so far?"

The clerk is given the luxury of a blank card called 'total'. But he still needs to know how to identify whether the sale was done in Mumbai. So we give him a card, where Mumbai has been written in the place allocated for 'location of sales'. He will take a card from his sales data, see the 'location of sales' column, compare it with the sample card we gave him; if the two match then he will note down the size of the sale into the 'total' card and then proceed to the next card. If the 'location of sales' column of the first card did not match Mumbai then he keeps it back and goes for the next card.

This is similar to making a salad by reading the recipe: 'take a stick of celery, clean it and cut it, go get the cucumber, clean and slice it, go get the…' and so on. You might say, "That is not a particularly intelligent way of making a salad!" Even a novice cook would read the list of materials needed to make the salad, clean them all up, peel them if necessary and then chop them and mix them with a dash of dressing. That in computerese, is called parallel processing. That is not how the vast majority of computers work. They do things one at a time, in a sequential way.

Of course, if he is really a dumb clerk how will he remember the procedure? So, for his sake, we have to write down the instructions or 'programme' in another place. The clerk now goes to the instruction file or the 'programme', reads the instruction and starts implementing it. When it is completed he goes to the next instruction. He reads it and executes the instruction and so on. The clerk will also need a scratch pad where he can do some arithmetic and wipe it out.

Thus, his 'memory' contains the programme instructions and sales data. The space where he compares the location of sale, adds the total etc form the 'processing unit'. He has a scratch pad or 'short-term memory', which he keeps erasing. Finally, he has a way of expressing the total sales in Mumbai on the 'total' card—the 'output' of all this hard work. He then waits for the next query to come from the eager beaver executive.

This is a simplified version of how a computer works but it has all the essentials. That is the reason why Feynman compares the computer to a filing clerk and a particularly stupid one at that.

You may have observed that there is nothing electronic or quantum physical about the computing process outlined. In fact, the theory of computing predates both electronics and quantum physics and evolved from seventeenth to twentieth century.

What differentiates man from rest of the species? His ability to make tools. That might mean, he is clever or lazy depending on the way you look at it. Just the same, all tools make particular tasks easier. Some tools extend our capabilities and reach new frontiers. Hunting weapons, flint stones, and practically all sorts of tools from the Stone Age till the twenty first century, have the two characteristics. One could in fact call man a technological animal. He fashioned tools for different tasks using the materials around him. Stone, wood, clay, bone, hide, metals, everything became raw materials for his tool making frenzy. Though a physically weak species, he asserted his domination through technology. Our understanding of why certain materials behave in a certain way or even how tools work in a particular way came with the development of speculative reasoning, experimentation and the scientific method. The empiricist came before the natural philosopher.

Computing started with the birth of numerals and arithmetic. Man began with mental sums and mathematical mnemonics like the sutras in Vedic Mathematics, inside the brain. Later, as computational load became too much to handle, man graduated to developing tools outside of his brain, to help him calculate the more complex or monotonous ones. Mesopotamians are credited with using sliding beads over wires for counting numbers around 3000 BC. Chinese improved on it two thousand years ago into the abacus. Even today many Chinese use the abacus for arithmetic. In fact an abacus competition held in 1991, in China, is reported to have attracted 2.4 million participants. In Europe, after the invention of Logarithms by John Napier in the seventeenth century, the Slide Rule was invented using his ideas. It became popular among engineers before the invention of electronic pocket calculators.

TERNARY CONVERGENCE

Computing machines have evolved over three centuries. French mathematician and physicist Blaise Pascal (1623-1662), created the first mechanical calculator in 1641, to help his father who was a tax collector.

He even sold one of these machines in 1645. It was remarkably similar to desktop mechanical calculators that were sold in the 1940s!

Gottfried Leibnitz (1646-1716), the great German mathematician, who invented differential calculus independent of Isaac Newton, dominated German science in the seventeenth century with his brilliance; much like Newton did in England. Leibnitz created a mechanical calculator called the Stepped Reckoner, which could not only add, but also multiply, divide and even extract square roots.

At that time the Indian decimal system of numbers, brought to Europe by Arab scholars, was dominating mathematics. It does so even today. A towering mathematician of France, Pierre Laplace (1749-1827), once exclaimed, "It is India that gave us the ingenious method of expressing all numbers by means of ten symbols, each symbol receiving a value of position as well as an absolute value: an important and profound idea, which appears so simple that we ignore its true merit." However, Leibnitz thought that a more 'natural' number system for computation is the binary system.

What is the binary system? Well, in the binary system of numbers there are only two digits, 0 and 1. All others are expressed as strings of 0s and 1s. If that sounds strange let us look at the decimal system that we are used to, where we have 10 numerals—0 to 9. We express all others as combinations of these. The position of the digit from right to left expresses the value of that digit. Thus, the number 129 is actually 9 units, plus 2 tens, plus 1 hundred. In symbolic terms 129= 1x10_2 +2 x10_1 +9 x10_0. If you were a mathematician, then you would say, "The value of the number is given by a polynomial in powers of ten and the number is represented by the coefficients in the polynomial."

In a similar way, we can write a number in powers of 2 as well. Thus the number 3 is 11, because 3 = 1 x2_1 + 1 x 2_0, and 4 is denoted by 100, since 4 = 1 x 2_2 + 0 x 2_1 + 0 x 2_0

.

ATROCIOUS ARITHMETIC: 1+1=0

Paul Baran invented the idea of packet switching which forms the basis of all modern networks, including the Internet. We will see his story in the chapter on the Internet later. But once, when he was asked why he did not study computer science, he said, "I really didn't understand how computers worked and heard about a course being given at the University of Pennsylvania. I was a week late and missed the first lesson. Nothing much is usually covered in the first session anyway, so I felt it okay to show up for the second lecture in Boolean algebra. Big mistake. The instructor went up to the blackboard and wrote 1+1=0. I looked around the room waiting for someone to correct his atrocious arithmetic. No one did. So I figured out that I may be missing something here, and didn't go back."

So let us see what Baran had missed, and why 1+1=0 made sense to the rest of the class.

Since there are only two digits in the binary system, we have to follow different rules of addition: 0+0=0, 1+0=0+1=1 but 1+1=0. The last operation leads to a 'carry' to the next position, much like if we add 1 and 9 we get a zero in the unit position but carry 1 to the next position. The real advantage of binary system appears during multiplication. In the decimal system, one needs to remember complex multiplication tables or add the same number several times. For example, if we have to multiply 23 with 79, then we use the multiplication tables of 9 and 7 and add the result of 23x9 to 23x70. Alternately, we can add 23 to itself 79 times. The result will be 1817.

In the binary system, however, 23 is represented by 10111, while 79 by 1001111. The multiplication of the two can be done by just 5 additions. Following the rules of binary arithmetic cited above, we get the answer: 11100011001, which translates to 1817. Voila! Without remembering complicated multiplication tables or carrying out 79 additions, we have got the result with just shift and carry operations.

There is nothing sacrosanct about the decimal system and one could do entire arithmetic with the binary system. Nevertheless, who cares for such convoluted games that reinvent good old arithmetic? Here came Leibnitz's insight that the *binary system is best suited for mechanical calculators*, which can follow the recipes: 0+0=0, 1+0=0+1=1 and 1+1=0 with
1 carried over!

Leibnitz had another great insight that *logic could be divorced from semantics and philosophy and married to mathematics*. He showed that instead of using language to prove or disprove a proposition, one could use symbols and abstract out the relation between symbols, into mathematical logic.

BOOLE AND HIS ALGEBRA

It took another two hundred years for another genius, George Boole, son of an English cobbler, a self-taught mathematician with no formal education, to connect up the two insights that Leibnitz had provided. He saw that the logical systems of Aristotle and Descartes, predominant in the West, posit only two answers for a proposition: true or false. He pointed out in a series of papers in the 1850s that such logic can be represented by the binary system. So he converted logical propositions and tests to prove or disprove them, to a prepositional calculus, using the binary system. His system came to be known as Boolean Algebra. Boole was not interested in computation but in logic.

It is interesting to note that not all logic systems are two-valued. There are several Eastern systems which are many-valued. In India there is a Jain school of logic called (*saptabhangi),* which is seven-valued, and a proposition can have seven possible outcomes! Similarly, the philosophy of (*anekantavad*) reconciles several points of view as being conditionally valid and posits that the fuller truth is a superposition of all.

BABBAGE GOES BALLISTIC

While Boole was interested in binary logic and theory, another English mathematician, with a practical bent of mind, Charles Babbage (1792- 1871), was actually interested in building a machine to carry out automatic computation. During those days, lengthy numerical calculations involving logarithms and other functions used to be made by a battery of human calculators and the results tabulated. The primary interest in these tables came from the artillery. The gunners wanted to know what should be the angle of elevation of the cannon to ensure that the shell would land on the desired target. A subject known as ballistics gives the answer.

A high school student might say, "Aha, that's easy. We have to solve the equations for projectile motion. The range of the shell will depend on the velocity of the ball leaving the cannon and the angle made by the cannon with the ground." That's true, but what we solve in schools is the idealised problem, where there is no air resistance. In real battles, however, there is wind, resistance of the atmosphere (which varies with the surrounding temperature and the height to which the shell is fired) and so on. After understanding these relationships quantitatively one ends up with complex non-linear differential equations. These equations have no simple analytical solutions; only hard numerical computation. Moreover, the gunner in the artillery is not a high-speed mathematician. He needs to crank up the turret and fire in a few seconds. At the most his buddy can look up a readymade table and tell him the angle. The more realistic the calculations are the more accurate can the gunner be in hitting his target. This was one of the driving forces behind the obsession with computation tables in the eighteenth and nineteenth centuries.

The preparation of artillery tables was not only labour intensive, but also highly error prone. The 'human computers' involved made several mistakes during calculations and even while copying the results from one table to another. The idea of a machine that could automatically compute different functions and print out the values looked very attractive to Charles Babbage. In 1821, he conceived of a machine called the Difference Engine, which could do this with gears and steps.

JACQUARD DOES IT

He started building the same, but then he came across an interesting innovation by a French engineer, Joseph Jacquard, which was used in the French textile industry. In 1801, Jacquard had devised a method of weaving different patterns of carpets using a card with holes that would control the warp and the weft as the designer desired. By changing the card in the loom, which came to be known as the Jacquard Loom, a different pattern could be woven. Voila, a textile engineer had invented the programmable loom!

Babbage saw a solution to a major problem in computation in Jacquard's punched cards. Till then all calculators had to be designed to calculate particular mathematical expressions or 'functions' as mathematicians call them. A major resetting of various gears and inner mechanisms of the machine was needed to compute a new function, but if the steps to be followed by the machines could be coded in a set of instructions and stored in an appropriate way as Jacquard had done, then by just changing the card one could compute a new function.

A LOVELY SOFTWARE ENGINEER

After twelve years of hard work, Babbage could not complete the construction of Difference Engine. In 1833, he abandoned it to start building what would now be called, "Version 2.0" that incorporated the programmable feature. He called it the Analytical Engine. Babbage's machine was based on the decimal system. His novel ideas attracted the young and lovely aristocrat Ada Lovelace, daughter of Lord Byron. With a passion for mathematics, she had numerous discussions with Charles Babbage and started writing programmes for the non-existent Analytical Engine. She invented the iterative 'Loop' and 'If,…Then' type of conditional branching. Modern day computer scientists recognize Lady Ada Lovelace as the first 'Software Engineer' and have even named a programming language as 'Ada'. Despite the lady's spirited advocacy and the hard work put in by Babbage and his engineering team, the Analytical Engine could not be completed. Finally, in 1991, to celebrate the second birth centenary of Babbage, the London Science Museum built the Analytical Engine.

SHANNON TIES IT ALL UP

The next major leap in the history of computing took place in late 1937, when Claude Shannon, then a Masters student at MIT, wrote his thesis on analysis of switching circuits. Vannevar Bush, a technology visionary in his own right, then ruled the roost at MIT. Bush had designed an analog computer to analyse electrical circuits called the Differential Analyser. The mechanical parts of the computer were controlled by a complex system of electrical relays. He needed a graduate student to maintain the machine and work towards a Master's degree.

Claude Shannon, a shy and brilliant student, signed up, since he liked tinkering with machines. As he started working on the machine, what intrigued him was the behaviour of relay circuits. The result was his thesis 'A Symbolic Analysis of Relay and Switching Circuits'. It is very rare that a Master's or a PhD thesis breaks new ground. Shannon's thesis, however, can claim to be the most influential Master's thesis to date, even though it took a couple of decades to be fully understood. He showed that circuits involving switches or relays, which go 'on-off' could be analysed using Boolean algebra. Conversely, since symbolic logic could be expressed in Boolean algebra, logic could be modelled using switching circuits. He showed that switching circuits could be used not only to carry out arithmetical operations, but also to decide logically: "if A, then B".

This was a remarkable insight. After all it is this ability to decide what to do, when certain conditions are satisfied while executing the programme,that distinguishes a modern computer from a desktop calculator. Ten years later, in a brilliant paper, 'A Mathematical Theory of Communication', Shannon laid the foundation of Information Theory that forms the backbone of modern telecommunications. However, in a 1987 interview to *Omni* magazine, he recalled, "Perhaps I had more fun doing that [Master's thesis] than anything else in my life, creatively speaking". The triangle of logic, binary computing and switching circuits came to be completed over a period of 300 years!

ALAN TURING: WHAT COMPUTERS CAN'T DO

Sometimes scientists worry about the limitations of an idea even before a prototype has been constructed. It seems to be a cartoonist's delight. Before the egg has even been laid, there is a raging discussion about whether the egg, when hatched, will yield a hen or a rooster. But that is the way science develops. Empiricists and constructionists—which most engineers are—like to build a thing and discover its properties and limitations but theoreticians want to explore the limitations of an idea 'in principle'. The theoreticians are not idle hair splitters. They can actually guide future designers on what is doable or not doable in a prototype. They can indicate the limits that one may approach, only asymptotically— always nearing it, but never reaching.

An English student of mathematics and logic at Cambridge, Alan Turing, approached computing from this angle in the mid-thirties and thereby became one of the founders of theoretical computer science. Turing discussed the limitations of an imaginary computer, now called Universal Turing Machine, and came to the conclusion that the machine cannot decide by itself whether a problem is solvable or whether a number is computable in a finite number of steps. If the problem is solvable then it will do so in a finite number of steps, but if it is not then it will keep at it till you actually turn it off.

Another mathematician, Alonzo Church, at Princeton University had preceded Turing in proving the same negative result on 'decidability' by formal methods of logic. The result has come to be known as the Church- Turing thesis.

However, in arriving at his negative result, Turing had broken down computing to a set of elementary operations that could be performed by a machine or Feynman's 'dumb clerk'. The

Turing machine was the exact mental image of a modern computer. Turing's contribution to computer science is considered so fundamental that Association for Computing Machinery (ACM), the premier professional organisation of computer scientists, has since 1966 instituted a prize in his name. The Turing Award is considered the most prestigious recognition in computer science and has assumed the status of a Nobel Prize.

All this feverish intellectual activity on both sides of the Atlantic, however, was punctuated by the rise of Nazism and emigration of a great number of scientists and mathematicians to America, and finally the Second World War.

GALILEO REVISITED: HOT WARS, COLD WARS AND COMPUTING

Human intellectual activity in the form of tool making technology or in the form of science—speculative reasoning combined with observation and experiment—has deep roots in curiosity and it has been amply rewarded by improvements in the quality of life. The history of mankind over ten thousand years is witness to that. However, destructive conflicts and wars have also been an ugly but necessary part of human history. Each war has used existing technologies and has gone on to create new ones as well.

Should scientists and technologists be pacifist savants of a borderless world, or flag waving jingoists? It is a complex question with only contextual answers, and only posterity can judge. Well-known German playwright, Bertolt Brecht, allegorised the relationship between scientists and war in his play *Galileo*. The famed scientist, hero of the play, is excited by the observations of celestial objects with his invention—telescope. Nevertheless, when the King (funding agency in modern terminology) questions him about his activities, Galileo says that his invention enables His Majesty's troops to see the enemy from afar!

The Second World War and the more recent Cold War have been classic cases where government funding was doled out in large amounts to scientists and technologists to create new defence related technologies. As we have noted earlier in the chapter on chips, electronics, microwave engineering and semiconductors were largely the result of the Radar project. One more beneficiary of the Second World War was computer science and computer engineering. Of course, with the Cold War anything that promised an edge against the 'enemy'—computers, communications, artificial intelligence—got unheard of funding.

In England computers were developed during the Second World War to help break German communication codes. Alan Turing was drafted into this programme. Across the Atlantic, computing projects were funded for pure number crunching to help the artillery shell the enemy accurately. Of course, there was the hush-hush Manhattan Project to create the 'mother of all bombs'. Physicists dominated the Atomic Bomb project, but thought it prudent to recruit one of the finest mathematical minds of the twentieth century, John von Neumann, a Hungarian immigrant. Foreseeing the horrendous amount of computation required in designing atomic weapons, von Neumann started taking a keen interest in the various computer projects across the US.

AN ARCHITECT CALLED JOHN VON NEUMANN

Von Neumann saw a promising computer in the Electronic Numerical Integrator and Calculator (ENIAC) at the University of Pennsylvania, and offered his advisory services. ENIAC, however, was a special purpose computer for artillery calculations and soon the idea to design and build a general purpose, electronic, programmable computer took shape. It was named Electronic Discrete Variable Automatic Computer (EDVAC). Von Neumann offered to write the design report for EDVAC, even though he was not directly involved in it. In between his preoccupation with the atomic bomb project at Los Alamos, he was intellectually excited by the challenges facing digital computing.

In his classic report, written in 1945, von Neumann decided to focus on the big picture, the abstract architecture of the computer—the overall structure, the role of each part and the interaction between parts—instead of the details of vacuum tubes and circuits. His architecture had five parts: Input, Output, Central Arithmetic Unit, Central Control and Memory. The Central Arithmetic Unit was a computer's own internal calculating machine that could add, multiply, divide, and extract square roots and so on in the binary system. The memory was the scratch pad where the data and the programme were stored, along with intermediate results and the final answer. Finally there was the Central Control unit which would decide what to do next, based on the programme stored in the memory.

But how would the Central Control Unit, together with the Arithmetic Unit, nowadays called the Central Processing Unit (CPU), go about executing the programme? Though his architecture was universal like Turing's imaginary machine, and he had various methods to choose from, Von Neumann chose what is now called scalar processing (*scala* in Latin means stairs, thus scalar implies a one-step-at-a-time process). He expected the central controller to go through an endless cycle: fetch the next set of data or instructions from the memory; execute the appropriate operation; send the results back to memory. Fetch, execute, send. Fetch, execute, send. If this reminds you of a particularly unintelligent way of making the salad, which we discussed earlier, then don't be surprised. It is exactly the same thing. Von Neumann chose it for the sake of reliability. To this day except for a few parallel computers that were designed after the late '70s, all computers are essentially based on this von Neumann architecture. In the '80s and '90s Seymour Cray's supercomputers followed a slightly modified version of this called the vector processing.

Von Neumann achieved two major things with his architecture. Firstly, by following the step-by-step method, he made sure that there was no need to worry about the machine's fundamental ability to compute. Like Turing's machine, it could compute everything that a human mathematician or any other computer could. This made sure that the hardware engineers could now worry about essential things like cost, speed, reliability, efficiency and so on and not whether the machine would work.

Secondly, with the concept of stored programme, he separated the procedure of solving a problem from the problem solver. The former is the Software and the latter the Hardware. The separation of the two was like separating music from the musical instrument. The same instrument, a sitar, a flute or a guitar, could be used to play a classical *raga* or hiphop-beebop.

Von Neumann's draft report on EDVAC in 1945 and another report on a new computer for the Institute of Advanced Study (IAS) at Princeton, written in 1946, shaped computer science for generations to come. They were a theoretical *tour de force.*

Von Neumann did not stop there. In the post-War period, amidst his hectic activities as a government advisor on defence technology and a fervent Cold Warrior, he kept coming back to various problems in computer science even though his mathematical interests were very wide. Among other things, he proposed the first Random Access Memory in 1946. His idea was that memory need not be stored and read only in a linear sequential fashion. For example, if you have a tape with music or a movie on it and if you want to see a particular scene then you have to forward it till you reach the spot. Where as a random access memory is like an index in a book, which lets you jump to a page where the keyword you were looking for appears. Implementation of this idea later led to tremendous speeding up of the computer, since sequential memory is decidedly slow.

In a series of reports, he laid the foundations of software engineering and also introduced new concepts such as the Flow Chart to show how the logic flows in a programme. He also pointed out that complex programmes could be built with smaller programmes, which are now called

Subroutines. The IAS computer was completed in 1952 and became a model for the first generation computers everywhere. Several replicas of this computer were immediately built at the University of Illinois, Rand Corporation and IBM. It was also replicated at the US defence laboratories at Los Alamos, Oakridge and Argonne, to help design the hydrogen bomb.

WERE COMPUTERS ONLY FOR ROCKET SCIENTISTS?

Other than highly specialised problems to be solved on an urgent basis like designing a nuclear bomb or cracking the enemy code, what are computers good for? Well, every branch of science and engineering needs to solve problems approximately using numerical methods in computers, since only a tiny set of problems can be solved exactly using analytical techniques.

Moreover, there are many cases where a complex set of equations needs to be solved fast, as in weather prediction, or flight control of a rocket. In some cases a large amount of data, like the census data or sales data of large corporations, needs to be stored and analysed in-depth to see trends that are not apparent. At times one needs to just automate a large amount of clerical calculation like payroll. These are just a few examples of calculations required in diverse walks of science, engineering and business where computers find application.

The full potential of this machine can be utilised, as we make the computer usable by people other than computer engineers. Consequently, a business can be made out of building computers for business applications. IBM was one of the first companies to realize it in the 1950s and invest in it. To this day, IBM has remained a giant in the computer industry.

How do you make the computer user-friendly? This is the question that has dogged the computer industry in the last fifty years. To the extent that computers have become user friendly, the market has expanded. And at a higher-level, human capabilities have been augmented.

The first step in this direction was the creation of a general-purpose computer. As software separated itself spirit-like from the body of the computer, it acquired its own dynamics as a field of investigation. Software engineers developed languages that made it easier to communicate with computers.

WHY CAN'T WE USE NATURAL LANGUAGES?

Hey, but why create new languages? We already have various languages that have evolved over thousands of years. Why not use them to instruct computers instead? Unfortunately, that is not possible even today. Natural languages have evolved to communicate thoughts, abstractions, emotions, descriptions and so on and not just instructions. As a result they have internalised the complexities and ambiguities of thought.

The human brain too has developed a remarkable capability to absorb a complex set of inputs: speech—emphasis, pause, pitch and so on; sight—hand movements, facial expressions; touch—a caress, a hug or a shove; and so on. We combine them all with our memory to create the overall context of communication.

Besides the formal content, the context helps to understand the intent of linguistic communication with very few errors. In fact, we are able to discern with a good bit of accuracy what an infant or a person not familiar with a language is trying to convey through ungrammatical utterances.

Despite all that we still spend so much time clearing up 'misunderstandings' amongst ourselves! Clearly, human communication is extremely complex and making a machine 'understand' the nuances of natural language has been well nigh impossible.

Look at some simple examples:

- The classic sitcom—a man hands a hammer to another and holds a nail on a piece of wood and says, "When I nod my head, hit it".
- The same word being used as a verb and a noun or an adverb and a verb as in—"Time flies like an arrow"and "Fruit flies like a banana".
- A modifier in the same position giving different implications— "The cast iron pump is broken" and "The water pump is broken".

Similarly, poetic expressions, like "Frailty, thy name is woman" or double *entendre*, "Is life worth living?—It depends on the liver", need an explanation even for human beings. An oft quoted result is the computer translation of "Spirit was willing but the flesh was weak" into Russian and back into English. The machine said, "Liqour was good but the meat was rotten"!

COMPUTERS NEED UNAMBIGUOUS INSTRUCTIONS

Instructions in natural languages can be full of ambiguity. For example, "Mary, go and call the cattle home" and "Mary, go and call the dogs home".

A computer programme containing a recipe on how to solve a problem needs to have very clear unambiguous instructions that Feynman's 'dumb clerk' inside the computer can understand. Hence the need for special programming languages, which look like English but are made up of words and a syntax, which can be used to give unambiguous instructions. But who would act as the 'grammar teacher' to correct the grammar of your composition, and whose word would be final?

A step in this direction was taken by Grace Hopper who wrote a programme in the early '50s that could take instructions in a higher-level language, more English-like, and translate it into machine code. The programme was called the 'compiler'. Since then compilers have evolved. They act as the 'grammar teacher' and tell you if the programme has been written grammatically correctly. Only then would the programme be executed. If not, it is 'detention' time for the programmer who needs to check the programme, instruction by instruction. This correction process is helped by error messages that say, "Line XYZ in the programme has PQR type of syntax error" and so on. The process is repeated till the programme is error free in terms of grammar.

Of course, the programmer can also make logical mistakes in his recipe, which then yield no result or an undesirable result. You are lucky if you can trace such logical errors in time, otherwise they become 'bugs' that make the whole programme 'sick' at a later time.

Apparently, the word 'bug' was first used in computing, when a giant first generation computer made of vacuum tubes and relays crashed. It was later discovered that a moth caught in an electro-mechanical relay had caused the crash! When IBM embarked on building a general-purpose computer, John Backus (Turing Award 1977) and his team at IBM developed a language—Formula Translation (FORTRAN) and released it in 1957. FORTRAN used more English-like instructions like 'Do' and 'If' and it could encode instructions for a wide variety of computers. The users could write programs in FORTRAN without any knowledge about the details of the machine architecture. The task of writing the compiler to translate FORTRAN for each type of computer was left to the manufacturers. Soon FORTRAN could be run on different computers and became immensely popular. Users could then concentrate on modeling the solution to their specific problem through a program in FORTRAN. As scientists and engineers learned the language there was a major expansion of computer usage for research projects.

Manufacturers soon realised that business applications of computers would proliferate if a specific language is written with business mode of data manipulation in mind. This led to the development of Common Business Oriented Language (COBOL) in 1960.

COMPUTER SCIENCE IN INDIA

It is interesting to see that computer science did not take very long to come to India. R Narasimhan was probably the first Indian to study computer science back in the late '40s and early '50s in the US. After his PhD in mathematics, he came back to India in 1954 and joined the Tata Institute of Fundamental Research, being built by Homi Bhabha.

India owes much of its scientific and technical base today to a handful of visionaries. Homi Bhabha was one of them. Though a theoretical physicist by training, Bhabha was truly technologically literate. He not only saw the need for India to acquire both theoretical and practical knowledge of atomic and nuclear physics, but also the emerging fields of electronics and computation. After starting research groups in fundamental physics and mathematics, at the newly born Tata Institute of Fundamental Research, Bhabha wanted to develop a digital computing group in the early fifties!

The audacity of this dream becomes apparent if one remembers that at that time von Neumann was barely laying the foundations of computer science and building a handful of computers in the USA. He had no dearth of dollars for any technology that promised the US a strategic edge over the Soviet Union. After the Soviet atomic test in 1949, hubris had been replaced by panic in the US government circles. And here was India emerging from the ravages of colonial rule and a bloody partition, struggling to stabilise the political situation and take the first the steps in building a modern industrial base. In terms of expertise in electronics in India, there did not exist anything more than a few radio engineers in the All India Radio and some manufacturers merely assembling valve radios. Truly, Bhabha and his colleagues must have appeared as incorrigible romantics.

Once he had decided on the big picture, Bhabha was a pragmatic problem solver. He recruited Narasimhan to TIFR in 1954 with the express mandate to build a digital computing group as part of a low profile Instrumentation Group. "After some preliminary efforts at building digital logic subassemblies, a decision was taken towards late 1954, to design and build a full scale general purpose electronic digital computer, using contemporary technology. The group consisted of six people, of which except I, none had been outside India. Moreover, none of us had ever *used* a computer much less designed or built one!" reminisces Narasimhan. The group built a pilot machine in less than two years, to prove their design concepts in logic circuits. Soon after the pilot machine became operational, in late 1956, work started on building a full scale machine, named TIFR Automatic Computer (TIFRAC), in 1957. Learning from the design details of the computer at the University of Illinois, TIFRAC was completed in two years. However, the lack of a suitably air-conditioned room, delayed the testing and commissioning of the machine by a year.

Comparing these efforts with the contemporary state of the art in the '50s, Narasimhan says, "Looking at the Princeton computer, IBM 701, and the two TIFR machines, it emerges that except for its size, the TIFR pilot machine was quite in pace with the state of the art in 1954. TIFRAC too was not very much behind the attempts elsewhere in 1957, but by the time it was commissioned in 1960, computer technology had surged ahead to the second generation. Only large scale manufacturers had the production know-how to build transistorized second generation computers."

TIFRAC, however, served the computing needs of the budding Indian computer scientists for four more years, working even double shift. The project created a nucleus of hardware designers and software programmers and spread computer consciousness among Indian researchers. Meanwhile, computing groups had sprung up in Kolkata at the Indian Statistical Institute and Jadavpur Engineering College as well. In the mid sixties TIFR acquired the first ever high-end machine produced by Control Data Corporation, CDC 3600, and established a national computing facility.

The third source of computer science in India came from the new set of Indian Institutes of Technology being established in Kharagpur, Kanpur, Bombay, Delhi and Madras. IIT Kanpur in particular, became the first engineering college to start a computer science group with a Masters and even a PhD programme in the late ’60s. That was really ambitious when only a handful of universities in the world had such programmes. H Kesavan, V Rajaraman at IIT Kanpur, played a key role in computer science education in India. “At the risk of not specialising in a subject, I purposefully chose different topics in computer science for my PhD students—thirty-two till the ’80s when I stopped doing active research— who then went on to work in different fields of computer science. Those days, I thought since computer science was in its infancy in India, we could ill afford narrow specialisation,” says Rajaraman. Moreover, generations of computer science students in India thank Rajaraman for his lucidly written, inexpensive textbooks. They went a long way in popularising computing. “At that time foreign text books had barely started appearing and yet proved to be expensive. So, I decided to write a range of textbooks. The condition I had put before my publisher was simple, that production should be of decent quality but the book be priced so that the cost of photocopying would be higher than buying the book,” he says.

The combination of research at TIFR, the Indian Institute of Science, Bangalore, and teaching combined with research at the five new IITs, led to a fairly rapid growth of computer science community. Today, computer science and engineering graduates from IITs and other engineering colleges are in great demand from prestigious universities and hi-tech corporations all over the world. However, Rajeev Motwani, an alumnus of IIT Kanpur, director of graduate studies and a professor of Computer Science at Stanford University, recalls, “I did my PhD work at Berkeley and am currently actively involved in teaching at Stanford, but the days I spent in IIT Kanpur are unforgettable. I would rate that programme, and the ambience created by teachers and classmates, etc, better than any I have seen elsewhere.” Winner of the prestigious Godel Prize in computer science and a technical advisor to the Internet search engine company Google, Motwani is no mean achiever himself Recently, on August 8, 2002, Manindra Agrawal, a faculty member at IIT Kanpur, and two undergraduate students, Neeraj Kayal and Nitin Saxena, hit the headlines of *The New York Times*—a rare happening for any group of scientists. Two days earlier they had announced in a research paper that they had solved the centuries old problem of a test for the prime nature of a number. Their algorithm showed that it is computable in a finite amount of time—in ‘polynomial time’ to be exact in computer science jargon. The claim was immediately checked worldwide and hailed as an important achievement in global computer science circles. Many looked at it as a tribute to education in IITs. “This is a sign of a very good educational system. It is truly stunning that this kind of work can be done with undergraduates,” says Balaji Prabhakar, a network theorist at Stanford University.

We will get back to the story of evolution of computers, but the point I am trying to make is that while we celebrate the current Indian achievements in IT, we cannot forget the visionaries and dedicated teachers who created the human infrastructure for India to leapfrog into modern day computer science.

TIME-SHARING: CIRCUMVENTING THE POOJARIS

Getting back to the initial days, the first generation computers made of valves were huge and expensive. For example, ENIAC when completed in 1946 stood 8 feet tall and weighed 30 tons. It cost 400,000 dollars (in the 1940s!) to complete. It had 17,468 valves of six different types, 10,000 capacitors, 70,000 resistors, 1,500 relays and 6,000 manual switches—an array of electronics so large that the heat produced by the computer had to be blown away using large industrial blowers. Even then, the computer room used to reach a temperature of forty-nine degrees Celsius!

The high cost of computers made sure that only a small set of people could be provided access. Moreover, even those few could not operate the computer themselves. That job was left to a handful of specialist operators. This meant that one could use the computer only as an oracle. The computer centre was the temple. One would go to it with his program in the form of a punched paper tape or a set of punched cards, hand it over to the *poojari*[†]—operator—and wait for the *prasaad*[‡]—the output. Of course, the oracle's verdict could be: "your offering is not acceptable", program has errors and the computer cannot understand.

If one wanted to change a parameter and see how the problem behaved then one would have to come back again with a new set of data. As von Neumann made extra efforts to introduce a diverse set of users to computing at Princeton, a new problem of too many users popped up— the problem of 'scalability'.

The users had to form a queue and the operator had to manage the queue. The easy way out was to ask everybody to come to the computer centre at an appointed time, hand in his or her program decks and then come back the next day to collect the output. This was known as 'batch processing', as a whole batch of programs were processed one after the other. This was most irritating. The second generation of computers was then built with the newly invented technology of transistors. This made the computers smaller, faster and less heavy, though they still occupied a room. But the problem of waiting for your output was not solved.

The solution to the problem was similar to a game that chess Grand Master, Vishwanathan Anand plays with lesser players. Anand plays with several of them at the same time, making his moves at each table and moving on to the next. Because he is very fast, he has been labelled as the 'lightening kid' in international chess circles. Hence, none of his opponents feels 'neglected'. Mostly they are still struggling by the time he comes back to them.

In computerese this was called Time Sharing, a technology that was developed in the late sixties. A central computer called the mainframe was connected to several Teletype machines through phone lines or dedicated cables. The Teletype machines were electronic typewriters that could convert the message typed by the user into electrical signals and send it to the mainframe through the phone line. They did not have any processing power themselves. Processing power even in the '60s and '70s meant a lot of transistor circuits and a lot of money. The computer, like Anand, would pay attention to one user for some time and then move to the next one and the next one and so on. It would allocate certain memory and processing power to each user. Because the computer played this game with great speed, the user could not detect the game. He felt he was getting the computer's undivided attention.

This was a great relief to the computing community. Now that one had interactive terminals, new programs were also developed to check the 'grammar' of the program line by line as one typed in. These were called Interpreters. Simultaneously new languages like BASIC (Beginners All-purpose Symbolic Instruction Code) were developed for interactive programming.

Bill Gates recalls in his book *The Road Ahead* the exhilaration he felt as a sixteen-year-old while using a time-sharing terminal along with Paul Allen in 1968 at the Lakeside School, Seattle. Gates was lucky. The Mothers Club at Lakeside had raised some money through a sale and with great foresight used the money to install a time-sharing terminal in the school, connected to a nearby GE Mark II computer running a version of BASIC. Gates says that it was interactivity that hooked him to computers.

[†] *Priest.*

[‡] *Offering blessed by the deity and returned to the worshipper.*

For most corporate applications of the ’60s and ’70s, like payroll and accounting, batch processing was fine. However, applications in science and engineering desperately needed some form of interactivity. Simulation of different scenarios and ‘what if’ type of questioning were an important part of engineering modelling. Thus time-sharing and remote terminals came as a form of liberation from batch processing for such users. Until the development of personal computers and computer networks in the ’70s and ’80s, time-sharing remained a major trend.

OF BITS AND BYTES AND FLOPS

By the way, before we go any further, we had better deal with some words and concepts that appear constantly in IT. The first is a ‘bit’. John Tukey and Claude Shannon at Bell Labs coined this in 1948, as a short form for a ‘binary digit’. The number of bits a logic circuit can handle at a time determines its computing power. An IBM team chose 8 bits as a unit and called it a ‘byte’ in 1957. It has become a convention to express processing power with bits and memory with bytes. The information flow in digital communication meanwhile is also expressed in bits per second (bps).

Greater the size of numbers that can be stored in the memory (also called ‘words’), more is the accuracy of the numerical output. Thus a powerful computer can process and store 64 bit ‘words’, whereas most PCs may use 16 or 32 bit ‘words’. The Central Processing Unit of the computer has a built-in-clock and the speed of the processor is expressed in million cycles per second or MHz. In the case of purely number crunching machines used for scientific and engineering applications, one measures the speed of the machine by the number of long additions that the machine can do accurately—called Floating Point Operations (FLOPs).

To make this bunch of definitions slightly more meaningful let us look at a few numbers. The first generation von Neumann’s computer at Princeton had a speed of 16 KHz. CDC 7600, a popular ‘supercomputer’ of the 70s had a speed of 36 Mega Hz. The Cray Y-MP supercomputer of the 90s had 166 MHz clock speed. On the other hand the IBM PC powered by Intel 8088 processor had a speed of 4.77 MHz and today’s Pentium-4 processors, sport Giga Hertz (a billion hertz) speeds.

Moreover, the Cray machine cost more than 10 million dollars whereas a Pentium PC costs about 1000-1500 dollars depending on the configuration! The incredible achievement is due to rapid development in semiconductor chip design and fabrication, which have been following Moore’s Law. Does it mean your PC has become more powerful than Cray Y-MP? Well not yet, because the Cray Y-MP could in a second do more than a billion additions of very long numbers and your PC cannot. Not yet.

DON’T FORGET MEMORY!

Other than the developments in the design and fabrication of CPUs, a concurrent key technological development has been that of memory. There are two kinds of memory, as we have seen earlier. One is the short term, ‘scratch pad’ like memory, which the CPU accesses while doing the calculations rapidly. This main memory has to be physically close to the CPU and be extremely fast as well. The speed of fetching and storing stuff in the main memory can become the limiting factor on the overall speed of the computer, as in the adage ‘a chain is as strong as its weakest link’. But since this kind of memory is rather expensive besides being volatile, one needs another kind of memory called the secondary memory, which is stable till rewritten, and much less expensive. It is used to store data and the programs. In the first generation days of von Neumann, vacuum tubes and oscilloscopes were being used for main memory and magnetic tapes were used to store data and programs. The first major development in computing technology along with the transistors was the invention of magnetic core memory by Jay Forrester at MIT in the early fifties. It was vastly more stable than the earlier versions. Soon the industry adopted it. But this memory too was expensive and a real breakthrough came about when IBM researcher Bob Dennard invented the semiconductor memory with a single transistor in 1966.

With the development of Integrated Circuit technology it then became possible to create inexpensive memory chips called Dynamic Random Access Memory (DRAM). A fledgling company in the Silicon Valley, called Intel, seized the idea and built a successful business out of it and so did a former supplier of geophysical equipment for the Oil and Gas industry in Texas called Texas Instruments. Both these companies are giants in the semiconductor industry today, with revenues of several billion dollars a year, and have moved on to other products (Intel to microprocessors and TI to communication chips).

The semiconductor DRAM technology has vastly evolved since the late ’60s. As for mass manufacture it has moved from 16 KB memory chips designed by Intel in 1968 to 4 MB in the mid-’80s at Texas Instruments and 64 MB in Japan in the late ’80s and 256MB in Korea in the late ’90s. One of the prime factors that made designing a personal desktop computer possible, back in the ’70s at the Palo Alto Research Center of Xerox, was the development of semiconductor memory. We will come back to the development of personal computing in the next chapter.

DIGITAL GODOWNS

The developments in the secondary memory, or what is now called storage technology, have been equally impressive. In the early days it used to be magnetic tapes. But, as we noted earlier, magnetic tapes are sequential. In order to reach a point in the middle of the tape, one has to go forward and backward through the rest of the tape. One cannot just jump in between. Imagine a book in the form of a long roll of paper and the effort to pick up where you left reading it. But book technology has evolved over the centuries. Hence, one has books with pages and even contents and index pages. It helps one to jump in between, randomly if need be. A computer memory of this sort was part of von Neumann’s wish list, in his report on the Princeton computer in 1946. But it took another ten years for it to materialise.

IBM introduced the world’s first magnetic hard disk for data storage in 1956. It offered unprecedented performance by permitting random access to any of the million characters distributed over both sides of fifty disks, each two feet in diameter. IBM’s first hard disk stored about 2,000 bits of data per square inch and cost about $10,000 per megabyte. IBM kept improving the technology to make the disk drives smaller and carry more and more data per square inch. By 1997, the cost of storing a megabyte had dropped to around ten cents. Thus while supercomputers of the early nineties had about 40 GB (billion bytes) of memory, now routinely PCs have hard disks of that capacity.

The great advantage of magnetic storage is that it can be easily erased and rewritten. If a disk is not erased, then it will ‘remember’ the magnetic flux patterns stored onto the medium for many years. This is what happens inside a tape recorder as well. The only difference is that in a normal audio tape recorder the voice signal is converted to smoothly varying ‘analog’ electrical signal, whereas here the signal is digital. It appears as tiny pulses indicating 1 or 0. The electromagnetic head accordingly gets magnetized and turns tiny magnets on the surface of the disk, ‘up’ or ‘down’ at a very high speed.

With increasing demand from banks, stock markets and corporations, which generate huge amounts of data every day, gigabytes of memory are being replaced by terabytes—trillion words. IBM Fellow at the Almaden Research Centre Jai Menon showed the author a mockup of a new storage system that would hold roughly 30 terabytes. The size of this memory was less than a two-foot cube. “By hugging this block, you will be hugging the entire contents of the Library of Congress in Washington DC, one of the largest libraries in the world,” pointed out Menon.

SUPERCOMPUTERS

The term ‘Supercomputer’ was popular in the ’80s and early ’90s but today no one uses it. Instead one uses ‘High Performance Computing’. The term ‘super’ or ‘high performance’ is relative and the norms keep changing. But such computers are primarily used in weather prediction, automobile and aeronautical design, oil and gas exploration, coding and decoding of communication for Intelligence purposes, nuclear weapon design and so on.

Many supercomputing enthusiasts even go to the extent of saying that they have a new tool for scientific investigation called simulation. Thus to design an aircraft wing to withstand extreme conditions or an automobile body to withstand various types of crashes one need not learn from trial and error but actually simulate the wind tunnel or the crash test in the computer and test the design. Only after the design is refined, the actual physical test needs be done, thereby saving valuable time and money. Since high performance computers are expensive, many countries, including India, have created supercomputing centres and users can log into them using high-speed communication links.

COMPUTERS AND BUSINESS

While number crunching is of paramount importance in science and engineering, the needs of business are totally different. Transactions and data analysis rule the roost there. The data can come from manufacturing or sales or even personnel. This led to the development of new tools and concepts of Databases and their Management.

Take transactions, for example. You are booking a railway ticket at a reservation counter in Mumbai for a particular train and at the same time a hundred others all over India also want a ticket for the same train and the same day. How to make sure that while one railway clerk allots seat number 47 in coach S-5 to you, the other clerks are not allotting the same to somebody else? This kind of problem is known as ‘concurrency’.

In pre-computer days, clerks maintained business databases in ledgers and registers. However, only one person could use the ledger at a time. That is why a familiar refrain in most offices of the old type was, “the file has not come to me”. But if many people can simultaneously share a single file in the computer, then the efficiency will improve greatly. But in that case one has to make sure that not everyone can see every file. Some might be allowed to only see it, while others can change the data in the file as well. Moreover, while one of them is changing the record in a file, others would not be allowed to do so, and so on. Software called ‘Database Management System’ takes care of these things.

A powerful concept in databases is that the data is at the centre and different software applications like payroll, personnel, etc use it. This view was championed by pioneers in databases like Charles Bachman (Turing Award 1973).

NEEDLE IN A HAYSTACK

As this concept evolved, new software had to be developed to manage data intelligently, serving up what each user wants. Moreover, since each application needed only certain aspects of this multidimensional elephant, called the data, methods had to be evolved to ask questions to the computer and get the appropriate answers. This led to the development of Relational Databases and Query Languages. A mathematician, Edgar Codd (Turing Award 1981), who used concepts from set theory to deal with this problem, played a key role in developing Relational Databases in the early seventies.

Now, if you are a large company engaged in manufacturing consumer goods or retailing through supermarkets, then as the sales data keeps pouring in from transactions, could you study the trends, in order to fine tune your supplies and inventory management or distribution to warehouses and retailers or even see what is 'hot' and what is not? "Yes, we can and that is how the concept of Data Warehousing emerged," says Jnaan Dash, who worked in databases for over three decades at IBM and Oracle. "While data keeps flowing in, lock it at some time and take the sum total of all transactions in the period we are interested in and try to analyse. Similarly, we can look for patterns in the available data. This pattern recognition or correlations in a large mass of data is called Data Mining", adds he. An example of data mining is what happens if you order a book from the Internet bookseller, Amazon. You will see that soon after you have placed the order, a page will pop up and say, 'people who bought this title also found the following books interesting', which encourages you to browse through their contents as well.

Ask the manager of a Udupi restaurant in Mumbai about trend analysis. You will be surprised to see how crucial it is for his business. His profitability might critically depend on it. It is a different matter that he does it all in his head. But if you are a large retail chain, or even a large manufacturer of consumer goods, then understanding what is selling where, and getting those goods to those places just in time, might make a big difference to your bottom line. Unsold stocks, goods returned to the manufacturer, the customer not finding what she wants and so on, can be disastrous in a highly competitive environment.

COMPUTERS AND FACTORIES

A major application of computers is in forecasting and planning within manufacturing. Let us say you are a large manufacturer of personal computers. You would like to use your assembly lines and component inventories optimally so that what is being manufactured is what the customers want and that when you have several models or configurations the machines produce the right batches at the right time. This would require stocking the right kind of components in the right quantities. In the old days when competition was not so severe in the business world, one could just stock up enough components of everything and also manufacture quantities of all models or configurations. But in today's highly competitive business environment one needs to do 'just in time manufacturing', a concept developed and popularised by Japanese auto manufacturers. It has been taken into the PC world by Dell, which keeps only a week's inventory! 'Just in time' lowers the cost of carrying unnecessary inventory of raw materials or components as well as finished goods.

However, this is easier said than done. "Even within a factory and a single assembly line some machines do their operations faster than the others. If this is not taken into account in production planning then an unnecessarily long time is taken up to make sure that a product will be ready at a particular time. Looking at these problems carefully, Sanjiv Sidhu developed a pioneering factory planning software in the early eighties", says Shridhar Mittal of I2 Technologies. "To achieve efficiency within several constraints and bottlenecks is the challenge", adds Mittal.

Sidhu strongly believes that with the appropriate use of IT and management of the whole supply chain of a company, manufacturing can be made at least fifty percent more efficient.

BREAKING OUT OF CONSTRAINTS

Talking of constraints, there are several very knotty problems in scheduling and optimisation. These have been solved by what is called Linear Programming. Mathematically the problem is reduced to several coupled equations with various constraints and programmes are written to solve these using well-known methods developed over two centuries. However, as the complexity of the problem grows even the best of the computers and the best of the algorithms cannot do it in reasonable time. There are several problems in linear programming that take exponentially longer time as the complexity grows. So even if one had the fastest computer, it would take forever to compute (more than the age of the universe) as the size of the problem increases.

In the early '80s, a young electrical engineer at Bell Labs, Narendra Karmarkar, was able to find a method, using highly complex mathematics, to speed up many problems in linear programming. His method is being applied very widely in airline scheduling, telephone network planning, risk management in finance and stock markets and so on. "It has taken almost fifteen years for the industry to apply what I did in three months and the potential is vast, because many problems in real life are reducible to linear programming problems", says Karmarkar.

MIMICKING THE MIND

What are the theoretical challenges in front of computer scientists? Creating Artificial Intelligence (AI) was one of them fifty years ago. Led by John McCarthy (Turing Award 1971), who also coined the word AI and Marvin Minsky (Turing Award 1969), several top computer scientists got together at Dartmouth College near Boston in 1956 and put forward the research agenda of Artificial Intelligence. Claims made at the Dartmouth Conference were: by 1970 a computer would

1. become a chess grand master;
2. discover significant mathematical theorems;
3. compose music of classical quality;
4. understand spoken language and provide translations.

Half a century, thousands of man-years of effort and millions of dollars of funding from the US Defence Department, have not got the AI fraternity nearer the goals. Except that a chess playing grand master was produced artificially in the 1990s.

Mimicking human intelligence has proved an impossible dream. Most early AI enthusiasts have given up their initial intellectual hubris. Raj Reddy, who was given the Turing Award in 1994 for his contributions to AI, says "Can AI do what humans can? Well, in some things AI can do better than humans and in many abilities we are nowhere near it. To match human cognitive abilities we do not even understand how they work".

If cognition itself has proved such a tough problem, then what about commonsense, creativity and consciousness? R. Narasimhan, another pioneer in AI, whose contribution to picture grammars and pattern recognition in the sixties is well known, says, "We do not even know how to pose the question of creativity and consciousness, much less of solving them".

Take computers understanding human languages, for example. The dream of a machine understanding your talk in Kannada and translating it seamlessly into German or Chinese has remained a dream. What we have is software to translate a very limited vocabulary or create a literal translation, which will require a human being to correct it and get the right meaning.

Rajeev Sangal and Vinay Chaitanya at IIIT (Indian Institute of Information Technology), Hyderabad are involved in such efforts to develop a package, "*anusaraka*," for various Indian languages. Using concepts developed by the great ancient Indian grammarian Panini, they are analysing Indian languages. Since all Indian languages form a linguistic group with many things common, they have found it easier to create Machine Assisted Translation packages from Kannada to Hindi, Telugu to Hindi and so on. This kind of work could help Indians speaking different languages understand each other's business communication better, if not each other's literature.

Meanwhile, many computer scientists, like Aravind Joshi at the University of Pennsylvania, one of the pioneers in natural language processing, are now studying how a two-year-old child acquires its first language! The hope is that it might give us some clues on how the brain learns complex and ambiguous language.

Neuroscientists like Mriganka Sur, head of the Brain and Cognitive Sciences department at MIT, are attacking the problem from another angle—that of understanding the brain itself better. Sur's work in understanding aspects of the visual cortex—part of the brain that processes signals from the optic nerve, leading to visual cognition—has been widely hailed. But this is just a beginning.

Understanding the brain is still far away. Theories that picture the brain as trillions of neurons that communicate in binary fashion are simplistic. They are unable to explain how a Sachin Tendulkar,[†] who has forty milliseconds before a Shoaib Akhtar's[‡] missile reaches his bat, is able to judge the pace, swing and pitch of the ball and then hit it for a sixer. The chemical communication system of neuro transmitters works at millisecond speeds and the neurons themselves work at a few thousand cycles' speed. Whereas, our gigahertz speed microchips with gigabit communication speeds can just about control a tottering robot and not recreate an Eknath Solkar,[1] a Mohammed Kaif[2] or a Jonty Rhodes[3].

"We currently do not understand the brain's cognitive processes. With all the existing mathematics, computer science, electrical engineering, neuroscience and psychology, we are still not able to ask the right questions", says Raj Reddy.

"We can write programs using artificial neural networks, etc to recognise the speaker, based on the physical characteristics of his voice, but we cannot understand human speech", says N Yegnanarayana at IIT Madras, who has done considerable work on speech recognition. "Purely physical analysis has severe limitations. After all, how does our brain distinguish somebody drumming a table from a Zakir Hussain[4] drumming it?" wonders Yegnanarayana.

WAITING FOR THE QUANTUM LEAP

While AI has been a disappointment, other challenges have cropped up. "Computer Science is a very young discipline, only fifty years old, unlike physics, which has been the oldest in scientific terms. Many times when new physical theories evolved they required inventing new areas of mathematics. One would expect that to understand what is computation and what can be done with it, there would be new areas of mathematics that will come up. The mathematical theory would help infer things prior to you doing it, then you may actually build a computer or write algorithms and

---

[†] The best known Indian batsman (cricket).

[‡] A reputed Pakistani pace bowler (cricket).

[1] Reputed Indian fielder (cricket).

[2] Reputed Indian fielder (cricket).

[3] Reputed South African fielder (cricket).

verify it. Constructing the mathematical theory of parallel computing is a challenge," says Karmarkar, winner of the prestigious Fulkerson Prize for his work in efficient algorithms.

Another person inspired by the science of algorithms is Umesh Vazirani, a young professor of computer science at Berkeley. Vazirani finds applying the concepts of quantum mechanics to computing very exciting. The area is called 'quantum computing'. In fact Vazirani proved an important result in 1992 that quantum computing can solve certain problems, which are intractable by today's 'classical' computers. Even though the example he chose to prove his thesis, was an academic one, it created a lot of excitement among computer scientists. After all the fundamental result in computer science has been Church-Turing thesis, which says that 'hard' problems remain 'hard' no matter what type of computer we use. Vazirani's work showed that quantum computers violated this basic belief.

Since then exciting results have been obtained in quantum algorithms that will have applications in the real world and hence there is a spurt of genuine interest and also unnecessary hype about the field. Peter Shor at Bell Labs proved one such result in the factoring problem in 1994.

What is the factoring problem? One can easily write down an algorithm to multiply any two numbers, however large, and a computer would be able to do it quite quickly. But if one takes up the reverse problem of finding the factors of a test number then the problem becomes intractable. As the test number grows bigger the time taken grows exponentially. For example, it is believed that a 250-digit number will take millions of years to factor, despite all the conceivable growth in computing power. Peter Shor showed that quantum algorithms could solve the factoring problem.

But what is the big deal? Why are millions of dollars being poured into theoretical and experimental research into quantum computing? Well, Shor's result shook up governments and financial institutions, because all encryption systems that they use are based on the fact that it is hard to factor a large number. Thus if somebody constructed a workable quantum computer of reasonable size, then the security of financial and intelligence systems in the world may be in danger of being breached!

The second interesting result has come from Lov Grover at Bell Labs, in 1998. Grover showed that quantum algorithms could be used to build highly efficient search methods. For example, it is as if you have a telephone book with a million entries and you wanted to search a name knowing the telephone number then there are a million unsorted pieces of data. This will take about half a million searches. However, Grover constructed a quantum algorithm to do the same in only 1,000 steps.

Others like Madhu Sudan at MIT, winner of the prestigious Nevanlina Prize for Information Sciences of the International Congress of Mathematics, think that the next big challenge to theory is to model the Internet."The standalone computer of Turing and Von Neumann has been pretty much studied intensively, but the network has not been and it might show some real surprises", says Madhu Sudan.

YEH IT, YT KYA HAI?

We started this chapter by talking about universality of the computer. But we have to understand it correctly and not hype it. Universality does not mean the computer can replace a human being or it is able to do all tasks that a human can. Universality means the computer can imitate all other

---

[4] A leading Tabla—a type of Indian drums, maestro.

machines. And except for a few mechanists of the seventeenth and eighteenth century, we all agree that human beings are not machines.

Moreover the computer cannot replace all other machines. It can only simulate them. Hence we have computer-controlled lathe, computercontrolled airplanes etc. So when we use catch words like the 'new economy' and so on, it definitely does not mean that bits and bytes are going to replace food, metals, fibres, medicines, buses and trains etc.

The hype about computers at the turn of the millennium led a colourful politician from Bihar to say, "*Yeh* IT, YT *kya hai*? (Why this hype about IT?) Will IT bring rain to the drought stricken?"

The answer is clearly 'No'.

Today we have great information gathering and processing power at our fingertips and we should intelligently use it to educate ourselves so that we can make better-informed decisions than before.

Computers cannot bring rain, but they can help us manage drought relief better.

FURTHER READING

1. Artificial Intelligence: How machines think—F. David Peat, Bean Books, 1988
2. Feynman Lectures on Computation—Richard Feynman, Perseus Publishing, 1999
3. The Dream Machine: J.C.R Licklider and the revolution that made computing personal—Mitchell M Waldrop, Viking Penguin, 2001
4. Men, machines, and ideas: An autobiographical essay—R. Narasimhan, *Current Science,* Vol 76, No 3, 10 February 1999.
5. Paths of innovators—R. Parthasarathy, East West Books (Madras) Pvt Ltd, 2000
6. Supercomputing and the transformation of science—William J. Kaufmann III and Larry L.Smarr, Scientific American Library, 1993
7. India and the computer: A study of planned development—C.R. Subramanian, Oxford University Press, 1992
8. Studies in the history of Indian philosophy Vol III—Ed Debiprasad Chattopadhyaya, K.P. Bagchi & Company, Calcutta, 1990
9. Supercomputers—V. Rajaraman, Wiley Eastern Ltd, 1993 10. Elements of computer science—Glyn Emery, Pitman Publishing Ltd, 1984
11. A polynomial time algorithm to test if a number is prime or not— *Resonance,* Nov 2002, Vol 7, Number 11, pp 77-79
12. Quantum Computing with molecules—Neil Gershenfeld and Isaac L. Chuang, *Scientific American,* 1998
13. A short introduction to quantum computation—A. Barenco, A. Eckert, a Sanpera and C. Machiavello, *La Recherche,* Nov 1996
14. A Dictionary of Computer—W.R. Spencer, CBS Publishers and Distributors, 1986
15. To Dream the possible dream—Raj Reddy, Turing Award Lecture, March 1, 1995 (http://delivery.acm.org/10.1145/240000/233436/p105-reddy.pdf?key1=233436&key2=4914509501&coll=GUIDE&dl=GUIDE&CFID=11111111&CFTOKEN=2222222)
16. Natural Language Processing: A Paninian Perspective—Akshar Bharati, Vineet Chaitanya and Rajeev Sangal, Prentice-Hall of India, 1999
17. Anusaaraka: Overcoming the language barrier in India—Akshar Bharati, Vineet Chaitanya, Amba P. Kulkarni, Rajeev Sangal and G. Umamaheshwar Rao, (To appear in "Anuvad: Approaches to Translation"—Ed Rukmini Bhaya Nair, Sage, 2002)

NIRVANA OF PERSONAL COMPUTING

"The fig tree is pollinated only by the insect Blastophaga Grossorum. The larva of the insect lives in the ovary of the fig tree, and there it gets its food. The tree and the insect are thus heavily interdependent; the tree cannot reproduce without the insect; the insect cannot eat without the tree; together they constitute not only a viable, but also a productive and thriving partnership. The purposes of this paper are to present the concept of and hopefully foster the development of mancomputer symbiosis."

— J.C.R. LICKLIDER, in his pioneering paper,
'Man-computer symbiosis' (1960).

"What is it [PC] good for?"

—GORDON MOORE, co-founder of Intel, recalling his response to the idea proposed by an Intel design team, of putting together a personal computer based on Intel's chips in 1975.

In the last chapter we saw glimpses of the power of computers to crunch numbers and their universal ability to simulate any machine. We also looked at some applications of this power. However, they were engineering or business applications, the plumbing of today's society, necessary and critical for running the system but hidden from most eyes. The users of those applications were a restricted set of academics, engineers, large corporations and government agencies.

Yet today there are nearly half a billion computers in the world and it is ordinary people who are using most of them. How did this happen and what new work habits and lifestyles have computers spawned? What new powers have they brought to our fingertips? How did machines that earlier intimidated users by their sheer size and price proliferate so widely and change our lives? These are some of the issues we will examine in this chapter.

GETTING PERSONAL

Instead of an apocryphal case study, let me narrate my own journey as a computer user. As a postgraduate student in physics, I took an undergraduate course in programming at IIT, Kanpur, more than thirty years ago. That was my introduction to computing and computers. The course looked interesting, but preoccupied as I was with classical and quantum physics, I looked at computers merely as an interesting new tool for numerically solving complicated equations in physics, especially when I could not find an exact solution through mathematical ingenuity.

What I did find was that computers were useful in dressing up my laboratory reports. A routine experiment, repeating the method devised by Robert Millikan (Nobel Prize for physics, 1923) to determine the charge of an electron, and another dealing with the vibrational spectra of diatomic molecules, needed some numerical calculations. I knew that the lab journal would look good if I added several pages of computer printout.

I used an IBM1620 at IIT, Kanpur's computer centre. It was an advanced computer for 1972-73, occupying an entire room and needing a deck-full of punched cards to understand my program written in FORTRAN. It took a whole day's wait to get the printout, and to find out whether there was an error in the work or the results were good. My achievement of using the computer for a routine lab experiment looked forbiddingly opaque and impressive to the examiner. He smelt a rat somewhere but could not figure out what, though he remarked that the experimental errors did not warrant a result correct to eight decimal places. He did not become privy to the cover-up I had done inside the lengthy computer printout, and I got an A in the course. Moral of the story: computer printouts look impressive. Numbers can hide more than they reveal.

My next brief encounter with computers, this time a time-sharing system at graduate school in Boston (1974-77), was perfunctory. My research problem did not involve numerical computing, since I was investigating the rarefied subject of 'super symmetry and quantum gravity in eight-dimensional space'. But for the first time I saw some of my colleagues use a clacking electric typewriter in a special room in the department, with a phone line and a coupler to connect them to the main computer, a CDC6600. They would type some commands and, seconds later, a brief response would manifest itself on the paper. For someone accustomed to daylong waits for printout, this appeared magical.

On my return to India, I took up another research problem, this one at IIT, Bombay (1978-80). It was a more down-to-earth problem in quantum physics, and needed some numerical calculations. My thesis advisor, an old-fashioned slide-rule-and-pencil man, depended on me to do the computations. Though there was a Russian mainframe at the IIT campus, I did my initial calculations on a DCM programmable calculator in the department. Having proved our hunches regarding the results, we needed a more powerful computing device.

We discovered a small Hewlett-Packard computer in a corner of the computer centre. It needed paper tape feed and had blinking lights to show the progress in computation. The BASIC interpreter, which had to be loaded from a paper tape after initialising the computer, made it interactive—the errors in the program showed up immediately and so did the result when the program was correct. We were overjoyed by this 'instant gratification' and higher accuracy in our computation. We went on to publish our research results in several international journals. Clearly, interactivity, however primitive, can do wonders when one is testing intuition through trial and error.

Ten years and some career switches later, I had become a writer and was glad to acquire an Indian clone of the IBM PC, powered by an Intel 286 microprocessor. It could help me write and edit. The word processing and desktop publishing function immediately endeared the PC to me, and continues to do so till today. I think the vast majority of computer users in the world today are with me on this.

In the early 1990s I was introduced to Axcess, a pioneering e-mail service in India, at *Business India*, where I worked as a journalist. It became a handy communication medium. Then we got access to the World Wide Web in the mid-'90s, thanks to Videsh Sanchar Nigam Ltd (VSNL), and a new window opened up to information, making my job as a business journalist both easy and hard. Easy, since I could be on par with any journalist in the world in terms of information access through the Internet. Hard, because the speed in information services suddenly shot up, increasing the pressure to produce high quality content in my stories before anyone else did and posted it on the Internet. The computer as a communication tool and an information appliance is another story, which we will deal with later.

The purpose of this rather long autobiographical note is to communicate the enormous changes in computing from central mainframes, to interactive systems, to personal computing. Older readers might empathise with me, recalling their own experience, while younger ones might chuckle at the Neanderthal characteristics of my story.

Nevertheless, it is a fact that the change has turned a vast majority of today's computers into information appliances.

SYMBIOTIC VISION

One of the visionaries who drove personal computing more than forty years ago was J.C.R. Licklider. Lick, as he was fondly called, was not a computer scientist at all, but a psycho-acoustics expert. He championed interactive computing relentlessly and created the ground for personal computing. In a classic 1960 paper, Man-Computer Symbiosis, Licklider wrote, "Living together in intimate association, or even close union, of two dissimilar organisms is called symbiosis. Present day computers are designed primarily to solve pre-formulated problems, or to process data according to predetermined procedures. All alternatives must be foreseen in advance. If an unforeseen alternative arises, the whole procedure comes to a halt.

"If the user can think his problem through in advance, symbiotic association with a computing machine is not necessary. However, many problems that can be thought through in advance are very difficult to think through in advance. They would be easier to solve and they can be solved faster, through an intuitively guided trial and error procedure in which the computer cooperated, showing flaws in the solution."

Licklider conducted an experiment on himself, which he quoted in the same paper. "About eighty-five per cent of my thinking time was spent getting into a position to think, to make a decision, to learn something I needed to know. Much more time went into finding or obtaining information than into digesting it. My thinking time was devoted mainly to activities that were essentially clerical or mechanical: searching, calculating, plotting, transforming, determining the logical or dynamic consequences of a set of assumptions or hypotheses, preparing the way for a decision or an insight. Moreover, my choices of what to attempt and what not to attempt were determined to an embarrassingly great extent by considerations of clerical feasibility, not intellectual capability. Cooperative interaction would greatly improve the thinking process."

Licklider left MIT to head the information processing technology office of the Advanced Research Projects Agency, ARPA, attached to the US defence department. He funded and brought together a computer science community in the US in the early 1960s. He also encouraged the development of computer science departments for the first time at Carnegie Mellon, MIT, Stanford and the University of California at Berkeley. "When I read Lick's paper in 1960, it greatly influenced my own thinking. This was it," says Bob Taylor, now retired to the woods of the San Francisco Bay Area. Taylor worked as Licklider's assistant at ARPA and brought computer networks into being for the first time, through the Arpanet. But that is another story, which we will tell later. For the time being it is important to note that after he left Arpa, Taylor was recruited by Xerox to set up the computing group at the Palo Alto Research Centre, the famous Xerox Parc.

THE SPARK AT PARC

One can safely say that in the 1970s Xerox Parc played the same role in personal computing that Bell Labs had played in the history of communications. Besides articulating ideas of interactive and personal computing, Parc pioneered the technology used by the Apple Macintosh, Microsoft Windows, as well as the laser printer – the point and click programs using the 'mouse' and layered windows. Parc also encouraged the design of graphic chips, which led to Silicon Graphics, and championed VLSI technology along with Carver Mead of Caltech. Small Talk, an object-oriented language that heavily influenced C++ and Java also originated there. The Ethernet was created at Parc to build a local area network and so was the Bravo word processor, which led to Microsoft Word.

No other group can claim to have contributed so much to the future of personal computing. Xerox made a couple of billion dollars from the laser printer technology invented at Parc, thereby more than recovering all the money it invested in open-ended research at the centre. However, as a document copier company faced with its own challenges, it could not win the PC battle. Xerox has been accused of "fumbling the future", but interested readers can get a well-researched and balanced account of Xerox's research in *Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age* by Michael Hiltzik.

The role played by Xerox PARC in personal computing shows once again that a spark is not enough to light a prairie fire: one needs dry grass and the wind too.

There are two important factors that fuelled the personal computing revolution. One is the well-recognised factor of hardware becoming cheaper and faster and the other the development of software.

MICRO IN SIZE BUT MEGA IN POWER

It is educative to note that the development of the microprocessor, which is the central processing unit of a computer on a single chip, was not the result of a well-sculpted corporate strategy or market research, but of serendipity. An Intel design team led by Ted Hoff and Federico Faggin had been asked by a Japanese calculator maker to develop a set of eight logic chips. Instead of designing a different chip for every calculator, the Intel team decided to design a microprocessor that could be programmed to exhibit different features and functions for several generations of calculators. The result was the world's first microprocessor— the Intel 4004.

While launching the 4004 in 1971, Intel pointed out in an ad in Electronic News that the new chip equalled the capacity of the original ENIAC computer though it cost less than $100. Gordon Moore went further and described the microprocessor as "one of the most revolutionary products in the history of mankind". A prescient declaration no doubt, but it was mostly dismissed as marketing hype. The world was not ready for it.

A year later Intel designed a more powerful 8-bit processor for a client interested in building monitors. The Intel-8008 led to glimmers of interest that it could be used inside business computers and for creating programmable industrial controls. But programming the 8008 was complicated; so very few engineers incorporated it in their industrial programmable controllers. But hobbyists thought it would be cool to use the chip to wire up their own computer, if they could.

Among these hobbyists were the sixteen-year-old Bill Gates and the nineteen-year-old Paul Allen. All their enthusiasm and ingenuity could not make the microprocessor support the BASIC programming language. So they instead made machines that could analyse traffic data for municipalities. They called their company Traf-O-Data. Several people were impressed by their machine's capability, but nobody bought any.

Then, two years later, Intel introduced the Intel–8080, ten times more powerful. That kindled immediate interest. A company named MITS announced the first desktop computer kit, called the Altair 8800, at a price of less than $400. Featured on the cover of Popular Electronics magazine's January 1975 issue, the Altair 8800 marked a historic moment in personal computing.

But, as Bill Gates and Paul Allen discovered, the machine did not have a keyboard or display or software, and could do no more than blink a few lights. True to form, the name Altair itself came from an episode of the then hugely popular sci-fi TV serial, Star Trek. Gates kick-started his software career by writing software to support BASIC on Altair computers. Later he took leave from Harvard College and, along with Paul Allen, started a microcomputer software company, Microsoft, at Albuquerque, New Mexico.

BIG BLUE AND THE PC

Intel then came out with the 16-bit microprocessor, the 8086, and a stripped down version, the 8088. At that time, 'Big Blue', that is, IBM, got into the act. Using Intel's processor and Microsoft's MS-DOS operating system, IBM introduced the PC in 1981. It took some time for the PC to start selling because enough software applications had to be written for it to be useful to customers, and a hard disk capable of holding a few megabytes of memory had still to be attached. This happened in 1983, when IBM introduced the PC-XT. Spreadsheets such as VisiCalc and Lotus 1-2-3, database management software like dBASE, and word processing packages like WordStar were written for it.

Meanwhile, a fledgling company, Apple Computers, introduced Apple Macintosh, which used a Motorola chip. The Macintosh became instantly popular because of its graphical user interface and its use of a mouse and multi-layered windows.

The PC caught up with it when the text based MS-DOS commands were replaced by a graphical operating system, Windows. Since then Windows has become the dominant operating system for personal computing.

What is an operating system? It is the software that comes into play after you 'boot' the PC into wakefulness. Few people today communicate directly with a computer. Instead, they communicate via an operating system. The more user-friendly the operating system the easier it is for people to use a computer. Although we talk of an operating system as if it is one single entity, it is actually a bunch of programs, which function as a harmonious whole, translating language, managing the hard disk with a provision to amend the data and programs on the disk, sending results to the monitor or printer, and so on.

Without an operating system a computer is not much more than an assembly of circuits.

A MULTIFACETED DEVICE

The significance of the proliferation of inexpensive peripheral devices, like dot matrix printers, inkjet printers, scanners, digital cameras, CD and DVD drives, and multimedia cards should never be underestimated. They have brought a rich functionality to the PC, moving it increasingly into homes.

Spreadsheets were the first innovation that caught the imagination of people, especially in accounts and bookkeeping. Bookkeepers and engineers have for long used the device of tabular writing to record accounts and technical data. While this was a convenient way to read the information afterwards, making changes in these tables was a tedious affair.

Imagine what happens when the interest rate changes in a quarter, and what a bank has to do. It has to re-calculate all the interest entries in its books. Earlier this took ages. Now an electronic spreadsheet allows you to define the relation between the cells of the various columns and rows. So, in the case of the bank above, a formula for calculating interest is written into the background of the interest column, when the interest rate is changed, the figures in the interest columns are automatically altered.

Imagine yourself as a payroll clerk preparing cheques and salary slips and the income tax rates change, or you are a purchase officer and the discounts given by suppliers change. The whole spreadsheet changes by itself with the modification of just one value. Isn't that magical?

FINANCIAL SECTOR DRIVES TECHNOLOGY

These techie machines made the accountant's life a lot easier. Computer use has since spread across the financial sector. Banks, stock exchanges, currency trading floors, investment bankers, commodity dealers, corporate finance departments and insurance companies have not only become early users of cutting-edge technology but even drive the creation of new technology. Database management, real-time event-driven systems, networking, encryption, transactions, disaster recovery, 24x7, aggregation, digital signatures, ‘publish and subscribe’ and so on are phrases that have come into software engineer’s jargon, thanks to demand from financial markets.

The dealing room of an investment banker today looks very similar to an air traffic control tower or a space launch command centre – with banks of monitors, high-end software and failsafe infrastructure. One can get a very readable and graphic account of this in *Power of Now* by Vivek Ranadive, whose event-driven technology has become part of the ‘plumbing’ across much of Wall Street.

Word processing and desktop publishing, or DTP, have come as a boon to everyone. After all, everybody writes, scratches off words and sentences, rewrites and prints. Isn’t it cool that you need not worry about your scribbled handwriting or an over-written letter that exposes the confusion in your mind? In a formal document, different fonts can separate the heading, sub-headings and text, and you can quickly incorporate charts, graphs and pictures as well.

These developments have had a much greater impact than the large office, and will continue to spread the benefits of computerisation.

THE BHASHA† EXPLOSION

If printing presses democratised knowledge to a great extent, then word processing and DTP have brought printing and publishing to our homes. The development of Indian language fonts and the software for DTP have given a remarkable boost to publishing and journalism in Indian languages.

“It was not money which drove us but the realisation that languages die if scripts die. If we want to retain and develop our rich cultural heritage then Indian language DTP is a necessity,” says Mohan Tambey, who passionately led the development of the graphic Indian script terminal, GIST, at IIT, Kanpur, during his M.Tech and later at C-DAC, Pune.

During the late ’80s and early ’90s, GIST cards powered Indian language DTP all over the country. Today software packages are available off the shelf. Some, like Baraha 5—a Windows-compatible Kannada word processor, are being freely distributed through the Internet. A tough nut to crack is developing Indian language spell-check programs and character recognition software. “The latter would greatly advance the work of creating digital libraries of Indian literature, both traditional and modern,” says Veeresh Badiger of Kannada University, Hampi, whose group is involved in researching ancient Kannada manuscripts. It is a non-trivial problem due to the complexities of compounded words in Indian languages.

RAJA RANI DEKHO†

Meanwhile, English language users can add a scanner to their PC and, by using optical character recognition software, digitise the text of a scanned document and build their own personal digital library. They can even clean a scanned image or fill it with different colours before storing it.

The capability to add multimedia features with a CD or DVD player has converted the PC into a video game console or an audio-video device, making it a fun gadget.

Instead of blackboards and flip charts, people are increasingly using PC-based multimedia presentations. The users are not just corporate executives, but teachers and students too. “Many people do not know that Power Point is a Microsoft product. It has become a verb, like Xerox,”

---

†An Indian language.

muses Vijay Vashee, the ex-Microsoft honcho who played a significant role in developing the product.

---

†*Form of rural entertainment for children with visuals of fascinating places and objects.*

A NOMAD'S COMPANION

The advent of laptops added a new dimension to personal computing— mobility. Though more expensive than desktop PCs, and used mainly by nomadic executives, laptops have become an integral part of corporate life.

To make presentations, work on documents and access the Internet when you are travelling, a laptop with an in-built modem is a must, preferably with a compatible mobile phone. In a country like the US, where there are very few 'cyber cafes', a travelling journalist or executive would be cut off from his e-mail if he did not have his laptop with him.

The major technical challenge in developing laptops has come from display and battery technologies. To create an inexpensive, high-resolution flat screen, is one of the main problems. "People all over are working on it," says Praveen Chaudhari a thin film solid-state physicist, at IBM's T.J. Watson Research Centre, Yorktown Heights. In 1995 Chaudhari won the National Technology Medal for his contribution to magneto optic storage technology, and was recently named director of the prestigious Brookhaven National Laboratory. His own work in developing the technology for large and inexpensive thin film displays might have a significant impact in this field.

"As for battery life, the benchmark for laptops in the US is 4.5 hours, since that is the coast-to-coast flight time," remarks Vivek Mehra, who played a key role in Apple's Newton II project. The Newton, a personal digital assistant, failed, but Mehra successfully applied what he learned there about consumer behaviour in the company he founded later: Cobalt Networks.

"In the case of all portables—laptops, PDAs or cellphones—a lightweight, powerful battery is the key," says Desh Deshpande, well known for his enterprises in optical networking. Deshpande is also the chairman of a company that is commercialising nanotechnology developed at MIT to produce better batteries.

In the mid-'90s, when multimedia applications began to be used extensively in desktops, here was a scramble to include these features in laptops. Prakash Agarwal, a chip designer, took up the challenge and designed a new chip, which combined the microprocessor logic with memory in a single chip. Memory and logic on a chip created magic and brought multimedia capabilities to laptops. Appropriately, Agarwal named his company Neomagic. At one time his chips powered about seventy per cent of the laptops in the world.

Designing chips that work at lower and lower voltages is another problem. "Lower voltages lead to lower power consumption and less heat generated by the chip, which needs to be dissipated. But this is easier said than done," says Sabeer Bhatia of Hotmail fame. Few people know that before he became the poster boy of Internet, Bhatia was working hard to reduce the voltages in chips at Stanford University.

IN YOUR PALMS

Not many people know that Sam Pitroda, whose name is associated with the Indian telecom revolution, is also the inventor of the digital diary, that handy gizmo which helps you store schedules, addresses, telephone numbers and e-mail addresses. Gradually digital diaries became more powerful and evolved into personal digital assistants, or PDAs. With models available at less than

$100, PDAs are fast proliferating among travellers and executives. They not only store addresses and appointments, they also contain digital scratch pads, and can access email through wireless Internet!

I came to know of another function of PDAs almost accidentally. An American software entrepreneur struck up a conversation with me as we waited outside Los Angeles airport. After picking my brain about the Indian software industry, he said at the end, "Shall we beam?" I had no idea what he was talking about. Turns out that 'beaming' is a new way of exchanging visiting cards. On returning from a conference or a business trip, it is a pain to input all the data from visiting cards into your computer or address book. The PDAs can store your digital visiting card and transmit the info at the touch of a button and using an infrared beam, to a nearby PDA.

No wonder a Neanderthal like me, using a dog-eared diary, was zapped. But that is what happens when the giant computers of von Neumann's days become consumer appliances. People invent newer and newer ways of using them.

WORKHORSES

Where PDAs and laptops constitute one end of personal computing, workstations constitute the other. Workstations are basically high-end PCs tuned to specialised use. For example, graphics workstations are used in computer graphics. They are also being used in feature-rich industrial design. For example, say, you would like to see how a particular concept car looks. The workstation can show it to you in a jiffy in three dimensions. Attached to a manufacturing system these workstations can convert the final designs to 'soft dies' in die making machines, to create prototype cars. The reason why the Engineering Research Centre at Tata Motors was able to launch its popular Indica quickly was that it used such applications to reduce the concept-to-commissioning cycle time.

As we saw earlier in the chapter on microchips, it is workstations that help engineers design chips.

High-end workstations can do computational fluid dynamics studies to help in aerodynamic design, as they did in the wing design of the Indian light combat aircraft, or are doing in the design of the civilian transport aircraft, Saras, at the National Aerospace Laboratory (NAL), Bangalore.

Roddam Narasimha, a distinguished expert in aerodynamics, took the lead in building a computer called Flo Solver, which could do complex computations in fluid dynamics at NAL. Of course, he did not use workstations; he built a parallel computer.

INDIAN AT THE OSCARS

Among the early users of graphics technology were advertising, TV and films. As a result, today's heroes battle dinosaurs in *Jurassic Park* and ride runaway meteors in *Armageddon*, or an antique on the table turns into an ogre and jumps at the *Young Sherlock Holmes*.

As Harish Mehta, one of the founders of Nasscom (the National Association of Software and Services Companies) puts it, "The Indian computer software industry should work closely with the entertainment industry to produce a major new thrust into animation and computer graphics".

Not many people know that during the *Star Wars* production in the 1970s, George Lucas, Hollywood's special effects genius, used some of the technology developed by an Indian academic-turned-entrepreneur, Bala Manian. Pixar, another well-known computer graphics company, also used a piece of Manian's technology of transferring digital images on to film—a technology he had developed in the '60s for use by medical experts looking at X-ray films. Manian was honoured for his contribution to Hollywood's computer graphics technology with a technical Oscar in 1999. "The screen in the auditorium showed a clip from *Adventures of Young Sherlock Holmes*, one of the many films that used my technology, as they announced my name," reminisces Manian, a shy academic with wide-ranging interests in optics, biomedical engineering and bio-informatics.

BREATHING LIFE INTO SILICON

Gordon Moore's question at the beginning of the chapter—"What is it [the PC] good for," when an Intel brain trust suggested in 1975 that the company build a PC—has to be understood in its context. Though Intel had the chips to put together a PC, without the requisite software it would have been a curiosity for electronics hobbyists, not a winner.

Today an increasing amount of software capable of diverse things has breathed life into silicon. Before the advent of the PC, there was hardly any software industry. The birth of the PC went hand in hand with the birth of the packaged software industry. If programming languages like BASIC, FORTRAN and COBOL hid the complexities of the mainframe from the programmer and made him concentrate on the modelling task at hand, packaged software created millions upon millions of consumers who employ the computer as an appliance to carry out a large number of tasks.

The complexities of programming, the complexities of the mathematical algorithms behind a word processing, image processing or graphical software are left to the software developers. A draftsman or a cartoonist need not worry about the Bezier curves or Spline functions involved in a graphics package; a photo-journalist downloading images from a digital camera into his PC for processing and transmission need not worry about coding theorems, data compression algorithms or Fourier transforms; a writer like me need not know about piece tables behind my word processor while editing.

SPIRALS

Let us step back a bit. Intel's failure to realise the opportunity in PCs, or Xerox's inability to commercialise the PC technology developed at its Palo Alto Research Centre under Bob Taylor's leadership, should be viewed with circumspection.

Every decision needs to be looked at in its historical context, not with 20:20 hindsight. In real life, the future is never obvious. In every decision there is an element of risk; if it succeeds others can look back and analyse what contributed to the success. But that does not guarantee a winning formula. Success is contextual, and the context is constantly changing.

Also there are the unknown parameters we call luck.

Bill Gates discovers positive and negative spirals in business successes and failures while analysing the super success of MS-DOS, Windows and Microsoft Office. The analysis shows that it is not the brilliance of one individual and his 'vision' that leads to success, but a host of factors acting together and propelling a trend forward.

Gates is sober enough to realise that he has been 'lucky' in the PC revolution and does not know whether he will be similarly successful in the Internet world, despite the tremendous resources, hard work, and focused research at Microsoft.

ECOSYSTEM OF A REVOLUTION

Clearly, all that we have discussed in this chapter shatters a popular romantic myth that long-haired school dropouts working out of garages in the Silicon Valley developed the PC. The PC triumph was the result of a vision carefully articulated by a number of outstanding psychologists, computer scientists, engineers and mathematicians supported by almost open-ended funding from the US defence department's Advanced Research Project Agency and from corporations such as Xerox, Intel and IBM.

The self-driven entrepreneurship of many individuals played a major role in the advancement of personal computing. To name a few prominent ones:

- Digital Equipment Corporation's Ken Olsen, who created the PDP minicomputer;
- Apple's Steve Jobs and Steve Wozniak;
- Microsoft's Bill Gates and Paul Allen;
- Jim Clarke of Silicon Graphics, famous for its computer graphics applications including animation and special effects;
- Andy Bechtolsheim, Bill Joy, Vinod Khosla and Scott McNealy of Sun Microsystems, which created workstations to fuel chip design and industrial design;
- John Warnock and Charles Geschke of Adobe Systems, who created software for desktop publishing, image processing and so on.

The contributions of several hardcore technologists cannot be ignored either:

- Wesley Clark, who developed the TX-2 at MIT, the first interactive small computer with a graphic display, in the 1950s;
- Douglas Engelbart (Turing Award 1997), who developed the mouse and graphical user interface at Stanford;
- Alan Kay, who spearheaded the development of overlaying windows, 'drag and drop' icons and 'point and click' technologiesto initiate action with his object oriented programming language—Smalltalk at Xerox Palo Alto research Centre;
- Carver Mead, who propounded VLSI (very large scale integrated circuit) technology at Caltech and Xerox Parc, which is today testing the physical limits of miniaturisation in electronics;
- Gary Kildall who created the first PC operating system, along with BIOS (basic input/output system);
- Dan Bricklin and Bob Frankston, with their first electronic spreadsheet VisiCalc; similarly, the inventors of WordStar and dBASE, which made the first PCs 'useful';
- Tim Paterson, the creator of MS-DOS;
- Mitch Kapor and Jonathan Sachs, with their spreadsheet Lotus 1- 2-3;
- Butler Lampson (Turing Award 1992) and Charles Simonyi with their word processing at Xerox Parc;
- Gary Starkweather, with his Laser Printer at Xerox Parc, to name a few.

Then there are the thousands who were part of software product development: writing code, testing programs, detecting bugs, and supporting customers. Similarly, hardware design and manufacturing teams came up with faster and better chips, and marketing teams spread the gospel of personal computing.

And let us not forget the thousands who tried out their ideas and failed.

I am not trying to run the credits, as at the end of a blockbuster movie. What I want to emphasise is that any real technology creation is a collective effort. The story of the PC revolution, when objectively written, is not pulp fiction with heroes and villains but a Tolstoyesque epic. It involves a global canvas, a time scale of half a century and thousands upon thousands of characters.

It is definitely not, as sometimes portrayed in the media, the romantic mythology of a few oracles spouting pearls of wisdom, or flamboyant whizkids making quick billions.

AND IT CONTINUES….

To illustrate the behind-the-scenes activity that fuels such a revolution, let me summarise a report from the February 2003 issue of the *IEEE Spectrum* magazine. Recently, about fifty software and hardware engineers and musicians from several companies shed their company identities and brainstormed for three days and nights at a Texas ranch. What were they trying to solve? The next quantum algorithm? A new computer architecture? A new development in nanotechnology that might extend the life of Moore's Law?

No. They had gathered to solve the problems of digital bits that translate into bongs, shrieks, beeps, honks and an occasional musical interlude when you boot your PC or when you blast a monster in a computer game. And they have been doing this Project BarbQ for the last six years!

A movement or a quantum leap in technology is the result of an ecosystem. Individual brilliance in technology or business acumen counts only within that context.

FURTHER READING

1. The Dream Machine: J.C.R Licklider and the revolution that made computing personal—M. Mitchell Waldrop, Viking 2001.
2. The Road Ahead—Bill Gates, Viking-Penguin, 1995.
3. Dealers of lightning: Xerox Parc and the dawn of the computer age— Michael Hiltzik, Harpers Collins, 1999.
4. Inside Intel: Andy Grove and the rise of the world's most powerful chip company—Tim Jackson, Dutton-Penguin, 1997.
5. Kannada manuscripts and software technology (In Kannada)—Dr Veeresh S. Badiger, Prasaranga Kannada University Hampi, 2000.
6. Digital Image processing—Rafael C Gonzalez and Richard E Woods, Addison Wesley, 1998.
7. Computer Graphics—Donald Hearn and M Pauline Baker, Prentice Hall of India, 1997.
8. Indians @ USA: How the hitech Indians won the Silicon Valley, *Business India* special issue, Jan 22-Feb 6, 2001.
9. Beyond Valuations—Shivanand Kanavi, *Business India*, Sept 17-30, 2001 (http://reflections-shivanand.blogspot.com/2007/08/tech-pioneers.html)
10. His Masters' Slave—Tapan Bhattacharya, CSIR Golden Jubilee Series, 1992.
11. The Power of Now—Vivek Ranadive, McGraw-Hill, 1999.

TELECOMMUNICATIONS: DEATH OF DISTANCE

"Electrical engineering is of two types: power engineering, which tries to send optimal amount of energy through a line, and communication engineering, which sends a trivial amount of energy through the wire but modulates it to send a meaningful signal"

—VANNEVAR BUSH,
a technology visionary, dean of MIT and
mentor of Claude Shannon.

Death of distance' is a catchy phrase to express a thought that mankind has had for long—"Oh, I wish I were there". Travelling in space and time has been right on top of the human fantasy list. It is as if the 'self', after becoming aware of itself, immediately wants liberation from the limitations and confines of the body, free to pursue its fancy.

We yearn to travel everywhere in space. Mundane economic reasons may have actually fuelled the transportation revolution: starting with the wheel and progressing to horse carriages, sail ships, steam locomotives, automobiles, bullet trains, airplanes and, finally, manned space exploration. But at the back of every transportation designer's mind is the dream of speed, the ability to cover vast distances in no time.

In this portion of the book, we will not talk about transportation or the teleportation of science fiction, but about an even more basic urge—the urge to communicate, the urge to reach out and share our thoughts and feelings. This is so basic a desire that it is at the foundation of all civilization and of society itself. Communication is the glue, the bond that builds communities. No communication, no community. Interestingly, the origin of both 'communicate' and community' lies in the Latin word '*communis'*— what is common or shareable. The urge to communicate created language.

The concepts of time and space travel would not have existed if somebody had not created imageries of different times and places—or history and geography. To access knowledge of a different time period we need to have access to archives of stored knowledge about that period, which is where history comes in. But to have access to another place we need quick transportation and instant communication.

Physical transportation has severe limitations imposed by the laws of physics; hence the ancient Indian *puranas*† talk of the mind's speed as being the swiftest. But if we can piggyback messages on faster physical carriers then we can achieve speed in communication.

These carriers could be the Athenian runners from Marathon, the poetic clouds in Kalidasa's *Meghadoot*,‡ carrier pigeons, or invisible waves of energy, like sound, electricity and light. The development of different carriers to convey our messages underlines the history of communications as a technology.

'Death of distance' is a dramatic way of expressing what has been achieved in this technology, but the driving force behind it is the basic urge of the human spirit to reach out.

ARE YOU LISTENING?

Voice dominates all human communication. There is a psychological reason for this. Voice and the image of a person add new dimensions to communication. The instrument that decodes the signal, our brain, perceives many more levels in a communication, when voice and image accompany language.

The auditory cortex and the visual cortex of our brain seem to trigger a whole lot of our memories and are able to absorb many more nuances than happens when we merely read a text message in a letter or e-mail, which is processed in the linguistic part of the brain. N. Yegnanarayana of IIT, Madras, who has spent over forty years in signal processing and speech technology, wonders how our brain is able to recognise a friend's voice on the telephone even if the friend has called after several years.

Data communication came into being with the advent of computers and the need to enhance their usefulness by linking them up. It has less than fifty years of history. Meanwhile digital technology has evolved, which converts even voice and pictures into digital data. This has led to data networks that carry everything—voice, text, music or video. This is called digital convergence.

The economic value of telecommunications services has resulted in large resources being deployed in research and development. This has led to rapid changes, and one of the crowning achievements of science and technology of the twentieth century is modern telecommunications technology.

An attempt to trace the history of telecommunications runs the risk of spinning out of control. There have been several forks and dead ends in the road traversed by communications technology, and these are closely associated with major developments in physics and mathematics.

Some historians have divided the history of telecom into four eras: telegraph and telephone (linked with electromagnetism); wireless (linked with electromagnetic waves); digital technology in signal processing, transmission and switching (linked with electronics); and optical networks (linked with quantum physics). I have employed the historical approach in some parts of this book, which I believe puts things in perspective and strips them of hype. But we will abandon it for a while and try a different tack. Let us start from the present and swing back and forth between the past and the future.

KNITTING AND COMMUNICATION

What is the telecommunications revolution all about? Consider these five broad themes:

- No man is an island: A personal communication system has come into being with cell phones, which allows us to reach a person rather than a telephone on a desk. With the spread of the mobile phone network over large parts of the urban world, instantaneous communication has become possible. At no point in time or space need one be out of reach. Well, that is a slight exaggeration in the sense that we cannot reach a person anywhere on the globe with an ordinary cell phone. But with global mobile personal communication systems (GMPCS) or satphones for short, we can be in touch with the rest of the world even if we are on Mount Everest. Satphones are not as affordable as mobile phones, so they are used only for special missions: mountaineering expeditions, media coverage of distant wars, on the high seas, etc. A geologist friend of mine, who was part of an Indian expedition to Antarctica, could keep in touch with his family in Baroda via satphone. The tag line in an ad published by a satphone service was: "The end of geography".

[†] Ancient Indian mythological texts.

[‡] *Messenger cloud*, Sanskrit poetic work by Kalidasa 4th century AD.

• Connecting India: Much before mobile satphones came into being, you could, thanks to communications satellites, use an STD, or subscriber trunk dialling, facility to call Agartala or Imphal in the north-east of India, or Port Blair in the Andaman Islands. Indian stock exchanges have built computer-based share bazaars in which one can sit anywhere in India—in Guwahati, Bhuj, Jullunder or Kumbhakonam—and do a trade. This is made possible by a network of thousands of VSATs (very small aperture satellite terminals). They can connect their trading terminals to the central computers through a satellite communications system. A network of automated teller machines, or ATMs, have spread around Indian cities in a short period of time, allowing for twenty-four-hour banking. You must have heard of the terrible cyclones that keep hitting the east coast of India, but have you heard of large-scale loss of life due to these cyclones in the last ten to fifteen years? No. The reason is the digital disaster warning system in the villages of the east coast, which is triggered by signals from a satellite. Also, satellites have made it possible for nearly a hundred channels to broadcast news, music, sports, science, wild life, natural wonders and movies in different languages. Twenty-four hours a day of information, entertainment and plain crap. (But that is not a technology issue).

• Telephony for the masses: In less than ten years from 1985 we got direct dialling facilities, or STD, to and from all major towns and cities in India and with voice quality greatly improved. The long wait outside a post office to make a ‘trunk’ call was replaced by instant connections at street corner STD booths. This was due to the conversion of the Indian telephone network into a digital system and the development of digital switching at the Centre for Development of Telematics, or C-DOT. The C-DOT class of switches, used in almost all exchanges in India, is one of the most widely used class of switches in the world. Long-distance communications have come to the doorstep of every Indian, despite a relatively small number of telephones per thousand of population.

• Global village: A World Wide Web of information and communication has come into being through the Internet, which has levelled the playing field for a student or a researcher in India vis-à-vis their counterparts in the most advanced countries. It has brought the dream of a universal digital library containing millions of books, articles and documents nearer to reality. Internet-related technologies have greatly reduced the cost of communication through e-mail, chat, instant messaging, and voice and video over Internet. They are making global collaboration possible. The Internet is also creating the conditions for a global marketplace for goods and services.

• Telecom for all: The cost of a long-distance telephone call has fallen steeply due to the digital communications revolution and rapidly falling bandwidth costs. Optical networking shaped this reality.

We will look at the evolution of the Internet in the next chapter, but let us now see how the other four features of the telecom revolution came into being.

PRE-HISTORY

The first major step in the telecom revolution was the digitisation of communications. But the difference it made will not be clear unless we get a broad picture of what existed before digital communications came into being.

It is worth noting that the first successful telecom device of the modern age was the telegraph, which, in essence, was a digital device that relayed a text message in dots and dashes of

Morse code. Almost as soon as electromagnetism was discovered in the 1800s, efforts began to apply it to communications. For almost a century from the 1830s the telegraph became the main long-distance communication medium.

The British colonial administration in India was quick to introduce telegraph to India. The first line was laid way back in 1851, between Calcutta and Diamond Harbour. After the 1857 uprising, the colonial government laid out a nationwide network of telegraphy with great alacrity, to facilitate trade, administration and troop movements.

The telegraph was also seen as a major development in international communications. A transatlantic submarine cable was laid way back in 1858. It took sixteen hours to transmit a ninety-word message from Queen Victoria—and then the cable collapsed. Lord Kelvin, an accomplished physicist and engineer (and founder of today's Cable & Wireless) took great pains to establish a viable transatlantic cable. A line connecting Mumbai and London, and another connecting Calcutta and London, were established in the 1860s.

But telegraph was quickly eclipsed by telephone. The magic of voice communication, pioneered by Alexander Graham Bell's telephone, was so overpowering that it quickly overwhelmed telegraph. "Watson, come here," Bell's humdrum summons to his assistant at the other end of his laboratory, became famous as the first words to be communicated electrically.

STRETCHING ANALOG TO ITS LIMITS

Financiers and entrepreneurs quickly saw the opportunity in telephony, and money poured into this service. The use of telephony increased dramatically in a short period of time. This kind of telephony was based on what communications engineers call 'analog' technology. What is analog? The word originates from the Greek word *analogos*, meaning proportionate. The electrical signal generated by your voice in the telephone speaker is proportional to it. As your voice varies continuously the signal also varies continuously and proportionately in voltage.

Several decades of engineering ingenuity led to an excellent voice communication system. As soon as vacuum tubes were invented, repeaters based on vacuum tube amplifiers were built in 1913. By 1915 a long-distance line was laid from the east coast of the US to the west coast— a distance of almost 3,000 miles.

With an increasing number of subscribers, pretty soon it became clear that the telephone company had to create a network. A network is actually a simple concept. We see networks in nature all the time. In our own body we see a network of nerves, a network of blood vessels, a network of air sacs in the lungs, and so on.

When many people have to be connected, connecting each telephone directly to every other would mean weaving a tangled web of cables. For example, to connect a community of 100 telephone users, 4,950 cables would have to be laid between them. The cost of the copper in the cables would make the system prohibitively expensive. An alternative is to create a hub and spoke structure, as in a bicycle wheel, so that all the telephones are connected to an exchange, which can connect the caller to the desired number. Such an arrangement brings down the number of cables needed to a mere hundred.

The system can be designed to optimise it for actual use and hence reduce costs further. Let's see how. If there are a hundred subscribers in a neighbourhood, then we find that on the average only ten per cent of subscribers use the telephone at one time. In a commercial complex the usage is higher but still far less than a hundred per cent. This fact allows the telephone company to optimise the size of the exchange.

TECHNOLOGY AND ECONOMICS

Cutting-edge technologies play a vital role in a mass service like telecommunications. Yet methods of organisation that shrink the service cost play an even more important role. The reach of the service and its economic viability to the service provider depends on managing costs while maintaining an acceptable quality of service.

That is why consumer behaviour studies play a major role in communications. Such studies enable the telephone company to build a network based on statistically verified real-life behaviour rather than a perfect network. "In a fundamental sense statistics play a vital role in the technology and economic viability of telecommunication networks, be they wired or wireless," says Ashok Jhunjhunwala of IIT, Madras, whose research group is doing innovative work in creating affordable telecom solutions in developing economies.

WHY DIGITAL?

Switching technologies developed from human switches—called 'operators'—to electromechanical ones. But problems of quality persisted in long-distance telephony, since the signals had to be amplified at intervals and the amplifiers could not filter out the noise in the signal. To overpower the noise in the signal, speakers would talk loudly during long distance calls. Today we can whisper into a telephone even when making an international call. We have the digitisation of communications to thank for that. Though vacuum tubes made digitisation technologically possible, the real digital revolution started with the invention of the transistor. Soon transistor circuits were devised to convert voice signals into digital pulses at one end and these pulses back into voice signals at the other end.

The basic idea behind digitising a signal, or what is now known as digital signal processing (DSP), is actually very simple. Suppose you want to send the picture of a smooth curve to a friend, how would you do it? One way would be to trace the curve at one end with an electrical pen, convert the motion of the pen into an electrical signal, which, at the receiver's end, is used to drive another pen that traces the same curve on a sheet of paper. This is analog communication.

Now think of another way. What if I send only a small number of points representing the position of the pen every tenth of a second. Since the pen is tracing a smooth curve and is not expected to jerk around, I send these positions to the receiving end as coordinates. If the receiver gets these points marked on graph paper, then by connecting them smoothly he can reasonably reproduce the original smooth curve.

Sounds familiar? Children do it all the time. They have these drawing books full of dots, which they have to connect to reproduce the original picture. If children knew it and drawing teachers knew it, then how come communication engineers did not?

The problem was: how many discrete points need to be sent to recover the original signal "faithfully"? To explain, let me extend the curve of the earlier example into a circle. Now, if you transmit the coordinates of four points on the circle, the way I did earlier, you may end up reproducing a quadrilateral rather than a circle at the other end. If you send data on more than four points, there is a good chance you will draw a polygon at the other end. As you send more points, you might approximate the circle quite well, approaching a situation, which mathematicians would explain thus: 'A polygon with infinite number of sides creates a circle.'

However, as we see time and again, *engineering is not about exact results but working results.* The crucial issue of how much to compromise and how pragmatic to be is decided by consumer acceptability and product safety. Here subjectivism, perceptions and economics all roll into one and produce successful technology.

Is the user happy with it, is it within his budget, can we give him better satisfaction at the same production cost or the same product with a lower cost of production? These are the questions that industry constantly battles with. The telecommunications industry, like other services, must choose appropriate technology that satisfies customers and does not drive them away by making the service too expensive. As the telecom industry worldwide has got de-monopolised, another issue has been added: can a telephone company give that little bit more in terms of features, quality and price than its rivals?

Let us now return to signal processing. The answer to signal recovery was provided way back in the 1920s by Henry Nyquist, a scientist at Bell Labs. He said that sampling the waveform at double the signal bandwidth would completely recover the signal. This is the first principle that is applied in digitising the voice signal. The voltage in the signal is measured 8,000 times a second, the resulting values are converted into binary zeroes and ones, and these are sent as pulses. In telecom jargon this is called pulse code modulation, or PCM. Pulse code modulation, achieved primarily at Bell Labs in the 1940s, was the first major advance in digital communications.

I did a sleight of hand there. We were discussing recovering the shape of the signal and suddenly brought in frequency components of a curve. The theory of 'Fourier transforms' allows us to do that. Jean-Baptiste- Joseph Fourier (1768-1830)—son of a tailor, a brilliant mathematician and an ardent participant in the French Revolution—pondered over complex questions of mathematical physics in the heat of the Egyptian desert while advising Napoleon Bonaparte. Today Fourier's theory is the bread and butter of communication engineers. And we leave it at that.

PIGGYBACK TO GLORY

Since noise and distortion were major hindrances in long-distance telephony, the application of digital signal processing led to dramatic improvement in quality. With a sufficient number of repeaters one could transmit voice flawlessly.

The next turning point came when the signal could also be transmitted in digital form. Human speech normally ranges in frequency from 300 Hz to 3,300 Hz. Coincidentally copper wires too transmit roughly this range of frequencies most efficiently with the least amount of dissipation and distortion. Of course, if you want to transmit a song by Lata Mangeshkar, the higher frequencies produced by the singer may get chopped. Hi-fi transmission is not possible in this mode but, if we use a very high frequency signal as a carrier and piggyback the voice signal on it, we can transmit a broad range of frequencies—also called bandwidth—thereby achieving even hi-fi transmission. This piggy-riding technology, invented by Major Armstrong, is called frequency modulation, and is similar to a walking commuter riding a high-speed train.

We could go a step further and transmit television signals, which need a thousand times more bandwidth along wires. That is what our neighbourhood *cablewallah* does. This kind of transmission requires coaxial cables—a copper core surrounded by a hollow, conducting cylinder separated by an insulator. (A simple copper wire will not do.) But high frequency signals are highly dissipated in cables, so how is this possible? This is where solid-state repeaters come into the picture. With their low power consumption, low price and compactness, we can insert as many repeaters as we need even at distances of a couple of miles or so in a cable.

SHARE AND SAVE

Telephony utilised this approach in a clever way. With the possibility of carrying high frequency signals, which allow a large range of frequencies or large bandwidth to piggyback on, engineers started multiplexing. Multiplexing is nothing but many signals sharing the same cable, like many cars using the same highway.

How do many cars use the same highway? They go behind one another in an orderly fashion (though not on Indian roads). But, on high-speed highways, if drivers try to overtake one another they can cause fatal accidents. In communications, too, there is something similar called 'data collision', which proves fatal to the two pieces of data that collide! There is need for lane discipline.

By the way, if you find me using a lot of analogies from transportation to explain concepts in telecommunication, don't be surprised. At a fundamental level, the two fields share many concepts and mathematical theories. Coming back to efficient ways of transportation, we can have double-decker buses where two single-decker busloads of people can travel at the same time, in the same lane, just by being at different levels. Communication engineers used the double-decker concept too. They sent signals at different frequencies at the same time through the same cable. With the high range of frequencies, or bandwidth, available on coaxial cables, this became eminently possible. Thus, if we have a cable with a bandwidth of 68 KHz, we can send sixteen separate voice channels of 4 KHz each. The remaining bandwidth of 4 KHz would be utilised for various signalling purposes. In engineering jargon this is called frequency division multiple access, or FDMA, which allows different signals to be filtered out at the end of the cable. This technology led to a steep improvement in the efficiency of cables, and it was soon adopted in intercity trunk lines.

Now consider yet another way in which efficiencies were improved. If you reserve a certain part of the available bandwidth—or 'channel', as it is also called—for one conversation, then you are being slightly extravagant. After all, even a breathless speaker's speech has pauses and, since a telephone conversation is usually a dialogue, the speaker at one end has to listen to the other party for roughly half the time. It has been found that a speaker uses the line for less than forty per cent of the time. Telecom engineers wondered if they could utilise the remaining sixty per cent of the time to send another signal, thereby easily doubling the capacity? They found they could, with a clever innovation called time division multiple access, or TDMA.

TDMA involves sending a small piece of signal A for a fixed time, then sending a piece of signal B and then again the next piece of signal A, and so on. The equipment at the other end separates the two streams of signals and feeds them to separate listeners as coherently as the speech of the speakers at the transmitting ends, and without any kind of cross-connection.

Only high-speed electronic circuits can allow engineers to split and transmit signals in this manner. Since our ears cannot discern the millisecond gaps in speech connected in this manner, it works out perfectly fine.

Remember Vishwanathan Anand playing lightning chess with several players simultaneously, an example I used to explain time sharing in mainframe computers in the chapter on computing? Well, TDMA is basically the same kind of thing. The same cable and same bandwidth are used to send several signals together. We could send pulses from one channel for a few microseconds, and then insert the next bunch from another signal, then the third and so on. Only a few milliseconds would have elapsed when we came back to the original signal.

For the last three decades, TDMA has been the preferred technology in telephone networks, but frequency division multiplexing—which had vanished with the disappearance of analog telephony—has come back with a bang in a different avatar. In state-of-the-art optical networks, a technology known as dense wave division multiplexing (DWDM) has allowed terabits of data to be sent down hair-thin optical fibres. With DWDM, laser signals are modulated with the data to be carried and several such signals of different wavelengths are shot down the same fibre. This is nothing but a form of frequency division multiplexing!

Good ideas rarely die; they reappear in a different form at a later date. After all, there are not that many good ideas!

The next thing that was needed was digitisation of the exchange, or the ‘switch’. This could be accomplished with the advent of integrated circuits and microprocessors. Thus the technology for digital communication was more or less ready in the mid-1970s, but to convert the networks to digital took almost twenty more years.

MATING CALLS AND MODEMS

Even today the copper wire between your telephone and the exchange— also called ‘the last mile’ or ‘local loop’—still uses mostly analog technology. Though ingenious new technologies called integrated services digital network (ISDN) and digital subscriber line (DSL) have come into being to make the last mile also digital, they are still to be deployed widely.

That is the reason why, when you need to connect your computer to the Internet, you need a modem. A modem converts the digital signals from the computer into analog signals and pushes them into the telephone line, and converts incoming signals from the Internet service provider into digital mode again so that your computer can process them. Hence the word modem, which stands for modulator-demodulator.

By the way, analog lines are not suited for data transmission; they are fine tuned to carry and reproduce voice well. They do that with devices like echo suppressors and loading coils. Unfortunately, these devices slow down data transmission. These obstacles can be overcome by sending a high frequency signal down the line to begin with.

Remember the screeching sounds that the modem makes when you dial up? They are nothing but signals generated by the modem and sent down the line to facilitate data transmission while simultaneously ‘talking’ to the modem at the other end for identification, or doing a ‘handshake’. A naturalist friend of mine, an ardent bird watcher, used to call them the mating calls of modems!

ADVANTAGE DIGITAL

What are the advantages of digital communications? There are many:

- Signal regeneration is done at every intermediate stage; this allows the system to tolerate a high level of noise, cross talk and distortion. It is rugged and unaffected by fluctuations in the medium.

- We can multiplex speech, data, fax, video, music, etc., all through one ‘pipe’.

- It allows for much higher data rates than does analog.

- One can control the network remotely.

- The signal can be very easily encrypted, allowing for greater privacy.

- The clinching factor: digital communications became more cost effective than analog with full digitisation of signal processing, transmission and switching.

MEN BEHIND THE SCENES

Let us look at some of the innovations in digital technology that have happened behind the scenes. When we, as consumers, see a good picture on the TV set, or get good voice quality on the telephone, we are scarcely aware of the thousands of innovators who have worked in the last fifty years to make these facilities possible.

Interestingly, as with computer science, the communications arena too has had several prominent Indian contributors. One reason for this is that some leading Indian educational institutions like the Indian Institute of Science (IISc), Bangalore, and IIT, Kharagpur, started teaching communication engineering quite early. "In the early 1960s, when digital signal processing was just evolving, IISc was perhaps one of the first institutions in the world to produce PhDs in this subject," says N. Yegnanarayana of IIT, Madras.

Teachers like B.S. Ramakrishna at IISc and G.S. Sanyal at IIT, Kharagpur, are remembered by hundreds of students who are now at the cutting edge of the telecom industry as researchers and managers. In the US professors, like Amar Bose at MIT, Tom Kailath at Stanford University, Sanjit Mitra at the University of California at Santa Barbara, P P Vaidyanathan at Caltech and Arogyasami Paul Raj at Stanford have carved out a unique place for themselves not only as first-rate researchers but also as excellent teachers.

INDIVIDUAL AND CORPORATE R&D

When we look for individual contributions, we have to be cautious. Says Arun Netravali, chief scientist, Lucent Technologies, "As technologies mature, it is more and more difficult for fully worked out great ideas to come from an individual. Innovations come increasingly from teams, in academia and industry. Teams are replacing the individual because the financial stakes in the communications industry today are high.

"The moment an innovative idea appears somewhere, venture capitalists and large companies are ready to invest millions of dollars into commercialising it. Time-to-market becomes crucial. It does not matter who lit the spark. Large teams are deployed immediately to develop the idea further and to commercialise it. Many Indians have made important contributions in the area of communications as members of large organisations; it is difficult to identify them as single sources of a technology."

Nevertheless, the contributions of several Indians stand out, and they have received public recognition for their work. Let us look at some of them.

ECHO…ECH…EC…E

One of the persistent problems in voice communications is echo. The problem could be solved in digital communication as Debasis Mitra, currently vice president, mathematical research, at Bell Labs, and Man Mohan Sondhi, now retired, showed back in the early seventies. Says Sondhi, "The echo cancelling equipment would have cost a prohibitive $1,500. It was only in 1980-81, as IC prices fell, that the echo canceller chip became economical and the telecom industry harnessed the technology. In fact, Donald Duttweiler, who made the chip, was also given the IEEE Award. It was one of the most complex special purpose chips at that time, and it had about 30,000 transistors. Today millions of echo cancelling chips are embedded in the network. In fact, there is a problem of echo in Internet telephony, in cell phones, in speakerphones and in teleconferencing equipment. So the echo cancellers are finding more and more applications."

The important point to note here is that *without the IC revolution the digital communications revolution could not have happened*. What has made the advances in communication theory and technology useful to the masses is the semiconductor industry churning out increasingly powerful chips at lower and lower costs, following the famous Moore's Law.

HOW MUCH IS TOO MUCH?

A major preoccupation of communications engineers is: What is the maximum capacity of a channel, and how much information can we push through it? And what is information itself? These fundamental questions bothered Claude Shannon at Bell Labs in the 1940s. He had joined Bell Labs after his epochal thesis at MIT connecting switching circuits with Boolean logic and laying the basis for digital computers (as we saw in the chapter on computing).

Shannon figured out the answers to his questions in communications theory in the mid-1940s, but he did not publish them till his boss pushed him. The result was the paper, *A Mathematical Theory of Communication*, published in two instalments in the *Bell Lab Technical Journal* in 1948.

That was the birth of information theory. Shannon's paper used mathematics too complex for the communications engineers of those times. It took some time for the impact to sink in; when it did, it was path breaking. A discussion on Shannon's theory is obviously beyond the scope of this book. Briefly, Shannon showed that communications systems could transmit data with zero error even in the presence of noise as long as the data rate was less than a limit called the channel capacity. The channel capacity, depends on bandwidth and the signal-to-noise ratio. The surprising result was that error-free transmission is possible even through a noisy channel. Though he did not show how to achieve maximum channel capacity, Shannon provided a limit for it.

Shannon's insights into the nature of information itself led to the whole field of coding theory and compression. Simply put, he argued that *real information in any communication is that which is unpredictable*. That is, if the receiver can guess what comes next then you need not send it at all! All compression techniques use this insight and see what is redundant. Then communication engineers go to great lengths to compress the signal through complex coding algorithms to push through as much information as possible through a given bandwidth.

SMS AND SHANNON

Actually this is not very different from what kids do nowadays, with SMS. They are driven by the restriction that they can send a maximum of 160 characters in a message. Thus, for example, when you send an SMS message to your friend: CU L8R F2F HVE GR8 WKND. These words might seem like gobbledegook to the uninitiated. However, your friend knows that you said, "See you later face to face. Have a great weekend". You have managed to send a 48-character message in 23 characters (including spaces). This is data compression. Millions of people who routinely use SMS may not know it, but they are using Shannon's information theory every day.

Interestingly, SMS has become so popular with youngsters today that the lingo is fast becoming a new dialect of the English language. The ultimate recognition of this has come from the venerable Concise Oxford English Dictionary itself, which has published a list of various acronyms frequently used in SMS and Internet chats, in its 2002 edition.

HOW LITTLE IS TOO LITTLE?

The name of the game in communications is optimising the use of bandwidth to reduce costs. One issue that has bothered engineers has been how to compress human speech and manage with much less than the 64 Kbits required for a toll-quality line.

Bishnu Atal found an innovative solution to this problem at Bell Labs in the 1970s. "Those days people did not take me seriously," recalls Atal. But his work was finally recognised in the '80s.

The question he asked was, "If we know the amplitude in speech in the past few milliseconds, can we reasonably predict its present value?" His answer was affirmative, and his solution became known as linear predictive coding.

Atal used his techniques for voice transmission at 16 Kbps and even 8 Kbps while maintaining a reasonable quality. The US military was immediately interested; as they saw that low bit rate communications was necessary in battlefield conditions. They also saw that with Atal's digital techniques encryption would become easy for secret communications.

"I did not want to work on secret projects, since that would have restricted my visits to India. So, I told them that I had done the required scientific work for compression and anybody else could work on encryption", says Atal. A Bell Labs fellow and a fellow of the National Academy of Engineering, Atal has now retired to teach at the University of Washington in Seattle.

With the advent of cellular telephony, where bandwidth is at a premium, any technique that can send voice or data at low bit rates is manna from heaven. A version of Atal's technique, called code excited linear prediction, is used in every cell phone today.

COMPRESS AND IMPRESS

While Atal worked to make low bandwidth channels useful for reasonable quality of voice transmission, there were others who were working on pushing high-bandwidth applications like high-quality audio and video through relatively lower bandwidths. Sending a full-motion video signal as it appears in the TV studio requires about 70 Mbps of bandwidth. Yet, amazingly, we can see a videoconference on a webcam or listen to MP3 audio, all on a simple dial-up Internet line of 56 Kbits. How is that? Scientists like Arun Netravali and N Jayant worked on techniques that made this possible.

Jayant and his team's work at the Signal Processing Research Lab at Bell Labs, related to audio transmission, led to the development of the 'MPEG Phase 1 Layer 3' (MP3) standard of audio compression. This technique was later commercialised by the Fraunhoffer Institute of Germany. Thanks to MP3 compression, we can now store hundreds of songs on an audio CD instead of the mere eight to ten songs we could earlier.

Netravali, currently chief scientist at Lucent Technologies, contributed enormously to digital video in the 1970s and 1980s. His work in video is widely recognised and used in media like DVD, video streaming and digital satellite TV. "In the 1970s and 1980s we had all the algorithms we needed, but the electronics we had was not fast enough to implement them," says Netravali. Then the microchip brought a sea change. It is good to see some of the technologies we worked on get commercialised." The Indian government honoured Netravali in 2001 with a Padma Bhushan, and the US government also in 2001, with the National Technology Award—the highest honour for a technologist in America.

Is compression a modern concept? No. That is how people have packed their baggage for centuries. Even your grandmother would say, "Keep the essentials and don't leave any free space."

WHAT PEOPLE WANT

This is not a politically correct sequel to Mel Gibson's movie, *What Women Want*, but an example of how perceptual studies have advanced communications. Netravali and Jayant's work is highly technical, but even laymen can understand some of the ideas used by them. They discovered that human perception, aural and visual, is remarkably inured to certain details. For example, Jayant and his team found that almost ninety per cent of the frequencies in high quality audio can be thrown away without affecting the audio quality, as perceived by listeners, because they get masked by the other ten per cent, and the human ear is none the wiser. This was great news for music companies, as they could now store hi-fi sound in a few megabytes of memory instead of a hundred megabytes.

Netravali also found that just applying coding algorithms would not provide enough compression to transmit full motion video. So he studied the physiology of the human eye and the cognitive powers of the viewer. What he found was this: if we are transmitting, say, the image of a person sitting on a lawn, then clearly we want good pictures of the person's face and body but not necessarily the details of the grass. Our eye and brain are not interested in the grass.

Similarly, when we transmit a head-and-shoulders shot of a person in motion, the motion makes only a small difference from frame to frame. What we need to do is calculate the speed with which different parts of the body are moving and estimate their position in the next frame, then subtract it from the actual signal to be sent in the next frame and instead send only the difference along with the coding algorithm. If we can do that, then we can achieve a lot of compression. Jayant and Netravali did this. They also studied the reaction of the eye to different colours and used the knowledge in coding colour information. The key factor in their approach was the analysis of perception.

What they did was not entirely new. In another context, Amar Bose, chairman of the Bose Corporation, applied psycho-acoustics to come up with his amazing speakers and audio wave-guides. "After my PhD at MIT in 1956, when I had a month's time before taking up a year's assignment in India, I bought a hi-fi set to listen to (I have always been a keen electronics hobbyist since my teens). But, to my horror, I found that despite having the right technical specifications, the equipment did not sound anywhere near high-fidelity," recounts Bose.

Bose then conducted psycho-acoustic experiments to see what people want when they hear music. He incorporated that information into the design of his equipment. As he continued to teach at MIT from 1958- 2001, his course in psychoacoustics was one of the most popular ones on the campus. Besides using good engineering, mathematics and digital electronics, Bose has continued to use psycho-acoustics as an essential ingredient in his products. The result: Bose is today rated as one of the biggest audio brands in the world.

The moral of the story: for a successful technology or a product, good engineering, mathematics and physics are not enough; we require perceptual inputs as well. We need a healthy mix of hard technology and 'what people want'!

SIR, YOUR TELEPHONE IS DEAD

While all these wonderful things were happening in the developed world, what was the state of telecommunications in India? The less said the better. Until the mid-1980s, the telephone was considered a rich man's toy, not an essential instrument for improving productivity and the quality of life. Though India produced top class engineers, they were migrating to the Bell Labs of the world, and the government, which had a monopoly over telecommunications, did not invest enough to spread the network and modernise it.

For their own colonialist reasons the British had introduced telegraphy and telephony to India quite early in the day, but independent India did not keep pace with the rest of the world. Even in the 1970s, Indian telegraphy was archaic (telex and teleprinters were just being introduced). The short forms and tricks employed by youngsters today, in SMS messages were then being used to save money by keeping telegrams short.

Curiously, while SMS sends your words faithfully there was no such guarantee with telegraphy. The mistakes made by telegraph employees sometimes caused avoidable heartbreak for job applicants and anxious relatives of seriously sick people.

The telephone service was notoriously bad. Telephones often turned up dead for days. Long distance telephony meant making a trip to the post office, "booking a call" and waiting for an hour or more for the operator to connect you. Often, after hours of waiting, the response would be: "The lines are busy". If you did get connected, the noise and distortion on the line were so bad that you had to not only shout at the top of your voice, throwing privacy to the winds, but also spend half the time repeating, "Hello, can you hear me?" This was India barely twenty years ago.

MOVING INTO THE TWENTY-FIRST CENTURY, FINALLY

Then things began changing. Today there are over eight lakh STD booths (public call offices) all over the country from which anyone can directly dial nearly 30,000 towns and cities in India and a large number of cities in most other countries. The networks have been expanded massively, with many more exchanges and it is so much easier to get a telephone connection. This has not only improved business communications but also communications among ordinary people, be it migrant labourers calling home or people keeping in touch with kith and kin. The poor use long-distance telephony as much as the upper classes and businesses.

Voice quality over telephone lines is excellent today; you don't have a problem getting a dial tone and the line you desire. The total number of telephones in India has increased from 5.5 million in 1991 to 30 million in 2002. As a result, the long waiting lists for telephones have vanished. You can get a connection practically on demand. New services such as mobile phones have been introduced, and already the number of mobile subscribers has crossed fifty million in ten years. Mobile phones are no longer associated with businessmen, stockbrokers, film stars and politicians. College students, taxi drivers, plumbers and electricians use them now. For artisans and other self-employed people, mobile phones seem to have become the much-needed contact points with customers.

Over three lakh route-kilometres of optical fibre have been laid in India in the last two decades. Broadband Internet services have become available to offices and homes, and their usage will grow as prices decline. But there is no space for complacency. China has achieved much more progress in telecom than India has, in the same period of time. China, too, had about 5.5 million telephones in 1991. In 2002 China had nearly 200 million mobile phones. China manufactures most of the telecom equipment it needs, including that required for optical networks within the country. We do have lessons to learn from the Chinese experience.

Fortunately, there is now widespread recognition in India that telecom is an essential component of infrastructure for the economic and social life of the country. The deregulation of the sector has led to large investments by many private sector companies. These new entrants are building their networks with state-of-the-art technology and providing the necessary element of competition by bringing in new and better services. Greater competition has resulted in a dramatic reduction in tariffs, expanding the market quickly. Clearly there is a sense of excitement in the air. Gartner Group has predicted that by 2007 there will be seventy million cell phone users in India. It may well surpass that.

If the past two decades have seen dramatic change, then the next twenty years may be hard to describe. The communication landscape of India will not be recognisable.

The achievement, in numbers, in the last twenty years is remarkable, and the progress in quality of service even more so. How did this transformation take place? Thousands of dedicated engineers and managers have endeavoured to change the scene, but there have been two key catalysts: the Indian Space Research Organisation (ISRO) and the Centre for Development of Telematics (C-DOT).

SOARING INTO THE SKIES

We will begin with communication satellites because they were the first hi-tech area to be developed and deployed to suit Indian requirements. Satellites have connected even the remotest areas of India through long distance telephony and national TV broadcasting. Today's buzz words such as distance learning, telemedicine and wide area networking of computers were first demonstrated and then implemented through Indian satellite systems in the 1970s.

The famous British science fiction writer, Arthur C Clarke, first mooted the idea of communication satellites. If you read Clarke's paper, *'Extra Terrestrial Relays—Can Rocket Stations give World-wide Radio Coverage*?', you would not believe, till you read the dateline, that it was written in 1945. At that time, there were no rockets (except a few leftover German V-2 ballistic missiles); and definitely no artificial satellites or space stations. But Clarke was audacious. He put forward a vision of three geo-stationary artificial satellites (see box on space jargon) hovering 36,000 km above the earth and being used as transmission towers in the sky to provide global communications coverage.

Many sci-fi aficionados believe that science fiction can be serious science, giving Clarke's satcom (satellite communication) as an example. As it happened, Clarke was a communications engineer who had worked on the radar project in the UK during the Second World War, and his paper is a serious scientific study, not a sci-fi story.

ABC OF SPACE TECHNOLOGY

Geostationary orbit

Any object placed in orbit at 36,000 km above the equator will take the same amount of time as Earth does to complete one revolution. This makes it stationary in relation to Earth. A dish antenna receiving signals from the satellite does not need to move to continuously track it, which makes tracking cheaper and less complex.

Transponder

A communication satellite used for telecom or TV receives the electromagnetic signal from the ground transmitter (uplink). It then retransmits it at a different frequency (downlink) towards Earth. The communication equipment on board a satellite that does the receiving and transmitting of such signals is called a transponder.

Why multi-stage rockets?

The heavier the weight that is carried into space, the larger must be the rocket ferrying it, because of the need for more fuel and power. It costs approximately $30,000 to put one kilo into geostationary orbit. In a multi-stage rocket the burnt out stages are detached one by one and drop to Earth so that less and less weight is actually carried into orbit.

Remote sensing

Observing Earth from a distance and getting information based on the reflective properties of different objects is known as remote sensing. Remote sensing can also be done using aircraft, but satellite remote sensing is far cheaper and more comprehensive.

What is digital direct-to-home broadcasting?

In DTH broadcasting, the signal frequency allows the consumer to receive the broadcast by means of a small dish antenna about a foot in diameter. Digital technology helps compress the signals so that many channels can be broadcast from a single transponder. This technology enables broadcasters to monitor and control usage since the signal can be keyed to individual users, who can then be charged subscription fees. Since it uses digital technology, DTH provides extremely high-quality picture and sound, as on a laser disc or CD. The satellite signals need to be decoded by a set-top box.

Why should we use liquid-fuelled rockets when solid-fuelled rockets are much simpler to make?

Solid-fuelled rockets cannot be turned on or off at will; once lit they burn till the propellant is exhausted. A liquid-fuelled rocket, on the other hand, can be easily controlled like the ignition key and accelerator of a car.

Arthur C. Clarke's dream of putting man-made satellites in space came true in the 1960s. The first step was the Russian Sputnik in 1957—a technology demonstrator rather than a communications satellite. It proved that a satellite could be injected into an orbit around the earth using rockets. The second big step was the launch of Telstar by AT&T in 1962 for a communications project. John Pierce, then president of Bell Labs, led the experiment. But Sputnik and Telstar were not geostationary satellites. They did not hover over the earth at the same spot, but zipped around every couple of hours. The first geostationary satellite was Syncom-2, launched by Hughes Aircraft Corp of the US in 1963. This made intercontinental TV broadcasting a reality. (It carried a live transmission of the Tokyo Olympics in 1964.)

In 1964, an international agency called Intelsat was created to provide satellite communications services. Intelsat launched the maiden international communications satellite, the Early Bird, in 1965. India was one of the first to join the Intelsat project, and has a place on the board of directors of the company. In fact, Videsh Sanchar Nigam Ltd (VSNL) is one of the largest shareholders in Intelsat, owning about five per cent of its equity. Intelsat has dozens of satellites in orbit over the equator above the Atlantic, Pacific and Indian Ocean regions, providing telephone and TV broadcasting services.

The concept of a communications satellite is actually quite simple. Radio waves have been used to send messages since the turn of the century. The Indian scientist Jagdish Chandra Bose was one of the pioneers in the field; he developed a range of microwave detectors for 12.5-60 GHz, made of iron and mercury. Bose's microwave coherer played a crucial role in the design of Marconi's wireless.

Certain ionised layers in the atmosphere reflect radio waves. This fact makes long distance communications, including radio broadcasts, possible. That's how a whole generation of us could listen to Amin Sayani on Radio Ceylon, and heard the Voice of America broadcasting a live commentary of Neil Armstrong's first steps on the moon and Radio Peking calling the peasant movement in Naxalbari as "the spring thunder over India".

TOWER IN THE SKY

TV broadcasting was possible only with much shorter wavelengths since only microwaves had enough bandwidth to carry the signal piggyback. The problem with short wavelength or high frequency waves is that they cover a small portion of the earth surrounding the antenna tower. In engineering terms this is called 'line of sight' communication. At any distance greater than sixty to eighty km the receiver will be 'invisible' to the antenna due to Earth's curvature. The only way to increase the reach of broadcast is to increase the height of the TV tower. That is why the tallest towers in the world—be it in Moscow, Toronto, New York or Paris—have TV antennae on them.

If the height of the tower is a vital factor in efficient broadcasting, then why not put the antenna up in the sky? That is the idea behind communications satellites. The only difference is that unlike TV towers, which originate the signal, satellites just relay the signal received from Earth back to Earth.

Earlier, short wave radio was used for intercontinental telephony too, making short waves bounce off the ionised layers surrounding Earth. But these layers constantly and unpredictably shift their characteristics; which is why intercontinental telephony was beset by a lot of noise. Satellites provided a great advantage since the relaying tower was not a dynamic ionised stratum of the atmosphere, but a reliable stationary satellite. Thus satellites became great platforms for intercontinental telephony as well.

The capacity of a submarine fibre optic cable, like the one connecting India to Dubai to Europe (Southeast Asia-Middle East-Western Europe) is many times that of satellites. Even so, satellites provide a low-cost alternative on certain routes. Over land they eliminate the need to dig trenches and bury expensive cable networks. To talk to Agartala from Mumbai you need a 'gateway' near Mumbai, a 'gateway' near Agartala and a satellite in the sky—and that's it. In fact, satellites proved for the first time that distance does not matter.

- A satellite call from Mumbai to Agartala, over 2,000 km away, costs the same as one to Pune, less than 200 km away.

- For nationwide TV broadcasting, instead of setting up a network of microwave towers every 30-50 km all across the landscape of India, which is a very expensive proposition, we can simply park a satellite in space.

- A satellite link can be set up in hours when needed. For example, after the Gujarat earthquake in 2001, the telecom network in Bhuj and nearby districts, including fibre optic cables, was damaged, but ISRO's satellite technology was immediately pressed into service to aid the administration in the quake-affected areas.

"A large part of basic satellite technology was developed at Comsat Labs in US and later at DCC and Hughes Network Systems," says Pradman Kaul, corporate senior vice president, Hughes Electronics Corporation and chairman and CEO of Hughes Network Systems. Kaul himself played a significant role in this.

As soon as satellite communications technology for TV broadcasting and telephony was developed, it became apparent that one of the irreplaceable uses of geostationary satellites would be mobile communications, where the transmitter and receiver are mobile, as in ships. Here, too, short wave radio was used for a long time, but sitcom provided reliable communication for the first time.

An international organisation, Inmarsat, was formed to provide maritime communications services in 1979. India, represented by VSNL is an early investor in Inmarsat, too. "In essence, Inmarsat provided the first global mobile phone service," says Jai P Singh, who himself played an important role first in ISRO and then in Inmarsat and ICO-Global. "Since the satellite is at a height of about 36,000 km above the earth, the terminal on the ship initially had to be bulky, with considerable transmitting power. However, Inmarsat then came out with a mobile phone that looks like a laptop computer, which can be used anywhere. Some of these instruments have high enough data rates to send pictures or flickering videos. Most journalists have been using this technology for newsgathering and transmission from remote areas of the world."

Several projects were initiated during the 1990s to make a lightweight satphone, available for telephony anywhere in the world. These included Iridium, with 66 low-orbit satellites, ICO-Global, with 12 medium-orbit satellites, and Globalstar, with 48 low-orbit satellites.

India, through VSNL, became one of the early investors in Iridium and ICO-Global. Both projects have faced financial problems. Iridium, promoted by Motorola, was the only project that was fully commissioned, but the high cost of the project ($7 billion), high user charges ($5-$10 per minute), and finally a very small number of subscribers led it to bankruptcy. Today, Iridium is being used mainly by the US department of defence. Globalstar, and New Ico Global, too, have gone through restructuring after near-bankruptcy. Recently, Globalstar has taken off in some parts of the world.

A geostationary satellite-based system called Thuraya is operating over West Asia. Most journalists in the recent US-led war against Iraq used satphones from Thuraya, which has been promoted by leading telecommunications companies in the UAE and other Arab countries.

The Indian foray into space and satellites started with an audacious dream way back in 1963. The dreamer was Vikram Sarabhai. Like Bhabha, Sarabhai too was a cosmic ray physicist. While Bhabha concentrated on developing Indian competency in nuclear energy, Sarabhai focused on applications of space technology when it was still in its infancy. Today, India has become one of the pre-eminent players in space technology.

The Indian Space Research Organisation has half a dozen advanced communications satellites of the Insat series in space. These were designed and fabricated in India. It has another half a dozen remote sensing satellites (the IRS series), making it part of an exclusive club of commercial remote sensing that counts the US and France among its members. It has its own rocket, the Polar Satellite Launch Vehicle (PSLV), which can launch a one-tonne satellite in a 400-1,000-km orbit for remote sensing purposes, and it is currently developing the Geostationary Satellite Launch Vehicle (GSLV) to launch a two-tonne communication satellite in the 36,000-km orbit above the equator. It has a state-of-the-art launch pad at Sriharikota near Chennai and its own master control facility to control satellites at Hassan in Karnataka. Indian technologists trained at ISRO have also contributed enormously to global satellite companies such as Intelsat, Inmarsat, ICO-Global, Panamsat, Loral and Hughes.

In just four decades, Indian space technology has come a long way at a fraction of the investments made by other countries. For example, the money invested in the entire Indian space programme from 1963 to 1997 was only half of the $2.4 billion Japan invested in developing its H-2 rocket in ten years. Yet the H-2, with a price tag of $150- 180 million per launch, was priced out of the market. Japan invested another $900 million to modify the rocket into H-2A, using all he manufacturing muscle of heavyweights like Mitsubishi, Kawasaki, Nissan and NEC to bring the launch cost to about $80 million. The H-2A has the same payload capacity as ISRO's GSLV, which is being developed at an additional cost of only $100 million to augment the capabilities of the earlier PSLV. No wonder, the prestigious aerospace magazine, *Aviation Weekly and Space Technology,* in a cover story in April 1997, hailed the Indian space programme as a "success with a shoestring budget".

THOSE MAGNIFICENT MEN AND THEIR FLYING MACHINES

It is hard to believe that it all started with a metre-long rocket a little bigger than a Diwali firecracker. How did ISRO leap to these heights from its humble beginnings? It took men of vision like Sarabhai and Satish Dhawan and thousands of innovative engineers and scientists to learn and improvise on technology that was at times not available through any foreign source at any price.

"It all started in a church where space is worshipped" might sound like a corny ad line or something from a sci-fi story, but it is a fact that the Indian space programme actually started in 1963 in a church and the adjoining bishop's house. While looking for a site to house the proposed Equatorial Rocket Launching Station, Sarabhai liked the spot, and the local Christian community at Thumba, near Thiruvananthapuram, graciously offered the premises for the cause. Scientists led by Sarabhai worked in the bishop's house, and the metre-long sounding rockets were actually assembled in the anteroom of the church and fired from a launch pad on the beach.

Pramod Kale, who retired as director of Vikram Sarabhai Space Centre at Thiruvananthapuram, remembers carrying out the traditional countdown for the first launch of a sounding rocket, at Thumba, to study the ionosphere. Today, the church, which has a history dating back to AD 1544, has been turned into the most comprehensive space museum in India, and has thousands of youthful 'worshippers' visiting every day.

In the 1960s, it was daydreaming of the highest order to think of an Indian rocket injecting an Indian satellite into orbit. But Sarabhai did just that and audaciously went ahead, realising his dream step by step.

"I was an undergraduate student studying physics when the Soviets launched the Sputnik. I made up my mind to join the space programme, though India did not have one then. Soon after my BSc honours, I went to Ahmedabad and met Dr Sarabhai. He asked me to come to Ahmedabad, finish my post-graduate studies and then join him in the Physical Research Laboratory," recalls Kale. He became one of the first to be roped into the space programme.

A characteristic feature of ISRO is its penchant for improvisation with whatever resources were available. Today, Indian remote sensing has come of age and its IRS data and expertise are in demand globally. Remote sensing as a technology was just emerging from the war-torn jungles of Indo-China, where it was deployed by the US to locate camouflaged Vietcong guerrilla positions. In the mid-1960s, when an opportunity came along to learn remote sensing in the US with the Earth Resources Technology Satellite project, Sarabhai grabbed it and sent Kale, P.R. Pisharoty, C. Dakshinamurthy and B. Krishnamurthy for the programme. These men later became well-known experts in the field.

"The first remote sensing experiment we did was driven by a very practical problem," says Kale. "There was this common scourge called 'coconut wilt' affecting coconut trees in Kerala. The disease affects the crown of the tree and cannot be seen from the ground, which means you can't estimate the damage. So we flew in a helicopter, and took pictures of coconut plantations using a camera with infrared-sensitive film. From that modest experiment, followed by decades of painstaking work, India has today become one of the global leaders in all aspects of remote sensing. This points to a defining characteristic of ISRO's work: it is driven by decidedly practical problems and inputs from a definite group of end users," says K. Kasturirangan, ISRO's former chairman.

ROCKETING AMBITIONS

Where rocket technology was concerned, the US refused to part with even the most elementary know-how because of the possibility of the technology being used to build missiles. They were only willing to sell their sounding rockets without any technology transfer. The French were more helpful. They sold solid-fuel technology for small sounding rockets. These were a far cry from the rockets required to launch satellites, but ISRO engineers like Brahm Prakash, Vasant Gowarikar and A.P.J. Abdul Kalam (now the president of India) led a focused effort to develop rocket technology.

ISRO went through a series of technology demonstrators like the SLV-3, ASLV and finally the now operational PSLV. The sophisticated, indigenously developed solid propellants in the first stage of the PSLV make it the third most powerful booster rocket in the world.

Solid-fuelled rockets are not enough to build an economical satellite launcher. To launch satellites, you need liquid-fuelled rockets, which are much more sophisticated. In the mid-1970s, France offered to share liquid propulsion technology in exchange for Indian collaboration in further development of the technology. ISRO engineers were to develop the pressure transducers for the Viking liquid engines then under development. While these transducers are hi-tech products, they are only a small component of the liquid-fuel engine. There are so many design complexities that 'know why' is absolutely essential to build an engine. The 'know how' in terms of drawings are not enough. Why does a component have to be machined to one-micron precision and not two micron? Why does one kind of gasket or 'O-ring' have to be used and not any other? Questions like these can make or break a rocket engine after millions of dollars of investment.

The French probably never expected Indians to learn the full technology. The contract was signed at a throwaway price.

'THESE INDIAN ARE CRAZY'

A fifty-strong team from ISRO worked in France between 1978 and 1980. It was made up of the cream of young ISRO engineers. Every day they brainstormed and sought solutions to complex design problems in the Viking. When they returned to India they asserted that they could build a sixty-tonne liquid engine. "We asked for only Rs 40 lakh to fund the project," recounted S Nambinarayanan, who led the team to France. "Prof Dhawan was crazy enough to humour us." Two years later, these engineers built a rocket engine model, and in 1984 they built an engine ready for testing. But India did not then have an adequate testing facility (built since then at Mahendragiri in Kerala); so the engine had to be taken all the way to France. The French engineers asked, "Is this your prototype? Do you have a manufacturing programme?" When the answer was in the negative, they could not believe it. According to Nambinarayanan, the French thought that the Indians were crazy.

The engine was tested and, to the jubilation of the Indians and the surprise of the French, it fired beautifully. Today's Vikas liquid engine used by ISRO is bigger than the French Viking engine and forms one of the essential workhorses of India's space chariots. Thereby hangs another tale of ISRO's ingenuity, improvisation and teamwork.

TECHNOLOGY DENIAL REGIMES

Launching a communications satellite weighing two tonnes or more requires even more powerful cryogenic engines, ones that use liquid oxygen and liquid hydrogen as fuel. The Russians were ready to sell the technology to India, and had even signed an agreement with ISRO in 1992; but the US invoked the Missile Technology Control Regime to bring pressure on Russia to withhold the technology. The Missile Technology Control Regime is an international agreement among ballistic missile owning nations, which aims to prevent missile technology from spreading to other countries.

But nobody on earth would think of using cryogenic engines for missiles since they need days of preparation. US policy did not make any sense, other than to pre-empt the emergence of a commercial rival. After all, launching communication satellites is a lucrative business and with

PSLV, ISRO had already shown that it could build one of the most cost effective rockets in the world.

The technology embargo could only delay ISRO by a few years. The agency bought six engines from Russia without transfer of technology and started building its own cryogenic engine, which would be ready in a few years. ISRO's track record makes its claim about developing cryogenic engines credible, even though they are an order of magnitude more complex than normal liquid-fuelled rockets. ISRO scientists are busy mastering the cryogenic technology at the Liquid Propulsion Systems Centre at Valiamala, Kerala.

PARALLEL PROCESSING

ISRO did not wait to develop a rocket system before mastering satellite technology. Like any ambitious organisation, it did some 'parallel processing'. The agency grabbed every opportunity that allowed it to gain experience. When the Soviet Union offered to carry an Indian satellite for free, ISRO quickly got down to designing and fabricating the first Indian experimental satellite, Aryabhata, named after the ancient Indian astronomer. The satellite was launched into a low earth orbit on 19 April 1975, and carried a scientific payload to study X-ray astronomy and solar physics.

Then came another generous offer from the Soviet Union to launch two satellites. ISRO designed and built the experimental remote sensing satellites Bhaskara-I and II, named after the ancient Indian mathematician. These were launched in 1979 and 1981, and gave ISRO some valuable experience.

Meanwhile Europe's Arianespace was trying to popularise its Ariane rocket, and offered to carry an Indian geostationary satellite free on an experimental flight, appropriately called the Ariane Passenger Pay Load Experiment (APPLE). ISRO immediately bit into it. "We worked feverishly to learn comsat technology from scratch," recalls U.R. Rao, former Chairman, ISRO.

Earlier, an opportunity had come up when the US offered its Application Technology Satellite-6 for an Indian experiment. Sarabhai immediately set his team into action. This led to the pioneering Satellite Instructional Television Experiment (SITE), the largest satellite based distance education experiment ever conducted. Under Yashpal's leadership, a team of engineers including E. Chitnis, P. Kale, R.M. Vasagam, P. Ramachandran and Kiran Karnik worked hard to make it a reality in 1975-76.

The earlier experience of building a satellite earth station at Arvi for the Overseas Communication Service (now VSNL) helped. Indian engineers also learned how to combine satellite signals with terrestrial low power transmitters to distribute TV in the local area. Eventually, this laid the basis for India's national TV broadcasting by Doordarshan during the 1982 Asian Games in Delhi.

The moral of the story is: ISRO is a success because of its pragmatic approach, its hunger to internalise new technologies available from others; and to dare to develop it indigenously if the technologies cannot be imported.

One of ISRO's life-saving innovations is its distributed disaster warning system. This system monitors weather pictures showing the progress of cyclone formations in the Bay of Bengal and broadcasts cyclone warnings via radio, TV and other means, including directly through loudspeakers in the villages on India's east coast. As a result, the number of cyclone-related deaths has declined since the 1980s.

An important aspect of India's space programme is its positive attitude towards transferring high technology to private manufacturers, helping them with technical upgradation as well as creating a nascent space industry. U.R. Rao, who took over from Satish Dhawan as chairman of

ISRO, worked hard to build a space industry in India by getting industrial vendors to produce components and sub-assemblies.

Today we have several companies such as L&T, Godrej, MTAR and Triveni Structurals as space-age equipment suppliers. After learning to manufacture to ISRO's extremely tough specifications and quality procedures, many suppliers found ISO 9000 and other such certifications child's play. As one wag put it, the documentation for a satellite weighs more than the satellite!

ISRO satellites have many other useful features, like search-and-rescue and global positioning. Today, not just long-distance telephony and TV but also ATM networks, stock exchanges, corporate data networks and even lotteries depend on the satellite systems.

G. Madhavan Nair, the current ISRO chairman, now has another audacious dream—that of reaching the moon. It looks like a daydream, but so did India's initial space programme seem in 1963 when Vikram Sarabhai launched a metre-long scientific rocket from the beaches of Thumba.

I think the point has been made sufficiently strongly that the first harbinger of the telecommunication revolution—in the broad sense of the term—in India was the space programme.

LET US SWITCH TO C-DOT

It's time we switched back to telecommunications. Let us get a glimpse of what happens when we make a telephone call.

When we lift the telephone handset, we almost immediately get a dial tone, then we press the keys on the dial pad to dial a number, and in a couple of seconds we get a ring tone (or a busy tone, in which case we decide to call back later). The person at the other end lifts his handset off the hook, and we talk. We end the call by re-placing the handset on the hook. This process is repeated with every call we make. At the end of the month, we get a bill for all the calls we have made, depending on the number of minutes we spoke for and the location we called (local or long distance).

We take this pattern for granted. We curse the telephone company if we do not get the dial tone, if the voice is not clear, if there are frequent disconnections, if there is cross-talk, or if there are mistakes in billing.

Now let us look inside the telephone network and see what actually happens when we make a call.

BEHIND THE SCENES

1. When the subscriber picks up her telephone, the switch, which scans the subscribers in its area every millionth of a second, detects that service is needed and the dial tone is transferred to that line. The mechanism then waits for the subscriber to dial.

2. The dialled number must now be used to set up a connection. The number is received as a train of frequency pairs from a push-button telephone. These signals cause the equipment to set up a path through the exchange to the appropriate outgoing line.

3. The line connecting the exchange to the receiver might be busy. It is necessary to detect a 'busy' (or 'engaged') condition and to notify the caller. Similarly, as there are only a limited number of paths through the exchange, the exchange itself may not be capable of making the connection. If the exchange is unable to make a connection, it will pass a busy signal to the caller's line. In a good network the latter would be a rarity.

4. The receiver's phone should then ring. Sending a signal down the line that activates the ringer does this.

5. The telephone of the receiving person is now ringing, but when that person answers, the ringing signal should be stopped. If nobody picks up the phone, the exchange may, after a respectable wait, disconnect the call.

6. When the call is successfully established and completed, and both the parties have put their telephones down, the circuit is disconnected, freeing the interconnection paths for other calls.

7. Last, there must be a way of recording the number of calls each subscriber makes and the duration and distance of long distance calls. This data is then used to produce month-end bills.

In the case of a long distance call, several exchanges and the trunk lines connecting them will be involved, and the process is slightly more complicated. But the essential point I am trying to make is that the exchange or the switch is the heart of the telephone network. Thus when a telecom system is to be modernised, one has to look at transmission and switching equipment. If transmission can be compared to the arteries and veins of the telecom body then the switch is clearly the brain.

We talked earlier of why switching is necessary for economical telephone networks; otherwise everybody has to be connected to everybody else. To make this cost saving, the telephone company must invest in building intelligence at the heart of the switching equipment.

In the early days, the most intelligent switches were used—human beings called operators. As telephone traffic increased, human beings proved inadequate to handle the rush, and new electromechanical relays and switches were invented to do the job.

Electromechanical equipment needed frequent maintenance as moving parts wore out very fast. The reliability of such equipment also decreased as traffic increased. Then transistors, and later integrated circuits and microprocessors, arrived on the scene as a *deus ex machina*.

The marriage of semiconductor technology and computing with telecommunications' switching needs led to the development of digital switches. These devices were essentially special purpose computers. The initial switches were mini-computers; only large metro exchanges could afford them. As microprocessors came into being and followed Moore's Law, the possibility arose of pervasive digital switching. Rapid adoption of digital switching in the 1970s facilitated better quality of service as well as lower costs.

There is another extremely important aspect of digital switching. Since the switch is actually a kind of computer whose capabilities are defined by the software written for it, whenever an upgrade is needed it can be achieved simply by writing new switching software.

In business, investments are not made only on the basis of the cost of equipment but what is called the 'lifecycle cost of service'. This includes the cost of the equipment, its maintenance, spares, consumables, and upgrading and support costs until the end of its designed life. At times equipment that is cheaper up front could mean larger costs over the life cycle. In the case of digital switching technology, we mainly need to upgrade the software, whereas previously an upgrade in electromechanical switches meant throwing out all the old switches and replacing them with new ones, which was, obviously, a time-consuming and costly process.

For a telecom company, digital electronic switching equipment has another important advantage over its analog predecessor: it uses microchips as its basic building blocks and therefore takes up little space. A large metropolitan switching station for 50,000 phone connections once occupied a six- to ten-floor building and needed hundreds of people to keep it operational. The same capacity can now be housed in one-tenth of the space and requires a staff of perhaps ten people to operate. The only serious drawback with the new technology is that digital switches produce heat and must be air-conditioned to prevent overheating. But that cost is small compared with the other costs that are eliminated.

THE COM IN INDIAN TELECOM

Sun Microsystems, the famous Silicon Valley computer maker, which sells a range of Internet servers, used to have an ad line a few years ago, which said: "We are the dot in .com". Obviously, the slogan was meant to advertise Sun's role in Internet infrastructure. If one were to coin a similar slogan for C-DOT, then it would be: "C-DOT is the com in Indian telecom".

Until the 1980s, Indian telecom was dominated entirely by electromechanical switches. This was one of the main reasons for bad telephone service. The Indian government was then looking at ways of modernising telecom. An obvious option was to import digital switches from the US, Japan or Europe. While this was the fastest route, there were primarily three drawbacks to it:

- India had meagre foreign exchange resources.

- The switches made by multinational companies were designed to handle a large number of lines (up to 100,000), and hence suited large cities. They did not have small switches that could handle about 100-200 lines, or the intermediate-range ones the country needed to spread telecom to small towns and large villages in India.

- It would have meant no incentive for indigenous R&D.

The question was, could India afford to spend enough money to develop its own switch and manufacture it at a competitive price? Even the most optimistic advocates of indigenous effort were sceptical, and they preferred to take the route of licensed production in agreement with a foreign multinational company. The CEO of a large multinational wrote to Prime Minister Indira Gandhi, cautioning her that his company had invested more than a billion dollars in developing the technology, and implying that it would be foolhardy for India to attempt to re-invent the wheel with its limited resources.

That accepted wisdom needed challenging. And the person who could dare to do so was Sam Pitroda, a Chicago-based telecom engineer from Orissa, who had studied in the US and participated in the development and evolution of digital switching technology. Pitroda had over thirty patents in the technology while working at GTE and later at Rockwell.
As an entrepreneur, he had done very well for himself financially.

In the early 1980s, he heard from a friend that Prime Minister Indira Gandhi had set up a high-level committee to look into the modernization of Indian telecommunications. He thought it was time he paid his dues to his country of origin. Having seen poverty and social discrimination in his childhood in his village, and now having become a participant in the worldwide IT revolution, Pitroda had no doubt that a modern telecom infrastructure would go a long way "in promoting openness, accessibility, accountability, connectivity, democracy, decentralisation—all the 'soft' qualities so essential to effective social, economic, and political development," as he wrote later in the *Harvard Business Review*.

Pitroda brought along with him his knowledge of technology, a 'can do' attitude and an impressive silvery mane he tossed while making a point, but not much else. He brought a breath of fresh air of optimism, aggression, confidence, flamboyance and media savvy into Indian telecom. He offered his services to the Indian government for one rupee a year.

And the offer was taken.

To recap the situation, in 1980, India had fewer than 2.5 million telephones, almost all of them in a handful of urban centres. In fact, seven per cent of the country's urban population had fifty-five per cent of the nation's telephones. The country had only twelve thousand public telephones for seven hundred million people, and ninety-seven per cent of India's six hundred thousand villages had no telephones at all.

"India, like most of the Third World, was using its foreign exchange to buy the West's abandoned technology and install obsolete equipment that doomed the poor to move like telecom snails where Europeans, Americans and Japanese were beginning to move like information greyhounds," asserts Pitroda in his characteristic fashion. "The technological disparity was getting bigger, not smaller. India and countries like her were falling farther and farther behind not just in the ability to chat with relatives or call the doctor but, much more critically, in the capacity to coordinate development activities, pursue scientific study, conduct business, operate markets, and participate more fully in the international community. I was perfectly certain that no large country entirely lacking an indigenous electronics industry could hope to compete economically in the coming century. *To survive*, India had to bring telecommunications to its towns and villages; *to thrive*, it had to do it with Indian talent and Indian technology", Pitroda added in his article.

Many discussions over three years, plus flying back and forth between New Delhi and Chicago, led to the establishment of C-DOT, the Centre for Development of Telematics. C-DOT was registered as a non-profit society funded by the government but enjoying complete autonomy. The Indian parliament agreed to allocate $36 million to C-DOT over 36 months to develop a digital switching system suited to the Indian network.

"We found five rooms in a rundown government hotel, and we went to work using beds as desks," says Pitroda of those early days. "A few months later, in October 1984, Mrs Gandhi was assassinated, and her son Rajiv became prime minister. He and I decided that I should press ahead with the initiative for all it was worth."

According to Pitroda, C-DOT engineers were conspicuously young, and they never seemed to sleep or rest. "C-DOT was much more than an engineering project. It did, of course, test the technical ability of our young engineers to design a whole family of digital switching systems and associated software suited to India's peculiar conditions. But it was also an exercise in national self-assurance. Years earlier, India's space and nuclear programmes had given the country pride in its scientific capability. Now C-DOT had the chance to resurrect that pride."

By 1987, within the three-year limit, C-DOT had delivered a 128-line rural exchange, a 128-line private automatic branch exchange for businesses, a small central exchange with a capacity of 512 lines, and was working on a 10,000-line exchange. The components for all these exchanges were interchangeable for maximum flexibility in design, installation and repairs, and all of it was being manufactured in India to international standards—a guaranteed maximum of one hour's downtime in twenty years of service! There was one problem; C-DOT had fallen short on one goal—the large urban exchange was behind schedule—but, overall, it had proved itself a colossal, resounding success.

What about the heat and dust in India and the need for air-conditioned rooms for digital switches? This was a serious issue for the country, large parts of which do not get a continuous supply of electricity. The solution was simple but ingenious. "First, to produce less heat, we used low-power microprocessors and other devices that made the exchanges work just slightly slower. Secondly, we spread out the circuitry to give it a little more opportunity to 'breathe'. The cabinet had to be sealed against dust, of course, but by making the whole assembly a little larger than necessary, we created an opportunity for heat to rise internally to the cabinet cover and dissipate," explains Pitroda.

The final product was a metal container about three feet by two feet by three feet, costing about $8,000, that required no air-conditioning and could be installed in a protected space somewhere in a village. It could switch phone calls more or less indefinitely in the heat and dust of an Indian summer as well as through the torrential Indian monsoon.

By November 2002, C-DOT switches equipped over 44,000 exchanges all over India. In the rural areas, ninety-one per cent of the telephone networks use C-DOT switches. Every village has not been covered yet, but we are getting there. Nationwide, 16 million lines, that is, forty per cent of the total operational lines in India, are covered by C-DOT switches.

Pitroda and Rajiv Gandhi also decided to open up the technology to the private sector. So C-DOT rapidly transferred the technology to over 680 manufacturers, who have supplied equipment worth Rs 7,230 crore and created 35,000 jobs in electronics. Seeing the ruggedness of these rural exchanges, many developing countries, such as Bhutan, Bangladesh, Vietnam, Ghana, Costa Rica, Ethiopia, Nepal, Tanzania, Nigeria, Uganda and Yemen decided to try them out.

For any institution, sustaining the initial zeal is hard once the immediate goals are achieved. After C-DOT's goals were achieved, the Indian telecom sector has gone through, and is still going through, a regulatory and technological upheaval. But that has not deterred C-DOT's engineers.

"It is creditable that through all this turbulence C-DOT has moved on to produce optical fibre transmission equipment, VSAT equipment, upgrading its switches to ISDN, intelligent networking, and even mobile switching technology. Today C-DOT may not be as high profile as it was in the 1980s, but it continues to provide essential hardware and software for Indian telecom despite intense competition from global vendors," says Bishnu Pradhan, a telecom expert who was among C-DOT leaders between 1990 and 1996.

THE STD BOOTH: A BRILLIANT SOLUTION FOR LOW TELEDENSITY

Before we move on to other parts of the communications revolution, let us note a characteristically Indian innovation not so much in technology as in management, which led to quantum leap in connectivity. That is the lowly public call office, or PCO, found at every street corner all over India today. These PCOs gave easy access to those who couldn't afford telephones, and brought subscriber trunk dialling to millions of Indians.

Public call offices are a part of any network anywhere in the world, so what is innovative about India's PCOs? The innovation lies in privately managed PCOs. As a result, we have over 600,000 small entrepreneurs running these booths and the telecom companies' income from long distance telephony has multiplied manifold.

The innovation also lies in realising that Indian society is essentially frugal in nature, and is amenable to sharing resources. What Pitroda did was to translate the Indian village and small town experience of sharing newspapers into the telecommunications scenario.

MOBILE FEVER

We now come to mobile phones, which have caught the world's fancy like nothing before. Today's communications world is divided into wired and wireless, denoting the way signals are exchanged. Wired communications have the great advantage of concentrating energy along a thin cylindrical strand of copper or silica. There is greater clarity in voice communications and much greater capacity to carry data. The disadvantage is obvious; you have to be available at the end of the wire to receive the message!

In wireless communications, the message rides piggyback on electromagnetic waves and leaves them in space (or ether as nineteenth century scientists called it). You can then reach any person, as long as he is in a position to receive those waves. He need not be at an office desk or at any fixed place where a wire can terminate. He can be almost anywhere on earth, provided certain conditions are fulfilled.

The caveats appear because buildings, trees, atmosphere, clouds and other such obstructions absorb electromagnetic waves. For example, you may not be able to receive a cellular call inside certain buildings. Some objects create 'shadows' so you may not receive the signal when you are behind them, for example, when you are in a valley or between tall buildings, or urban canyons. Then there is the effect of the earth's curvature, which compels you to be in the line of sight of a transmitter or repeater. But, despite all these problems, it has become possible for people to talk to one another, regardless of where they are.

The change in the communications culture brought about by wireless cellular phones can be seen by the simple fact that the opening of a conversation is no longer, "Hello, who is this?" but "Hello, where are you?"

The weak link in wireless communications is that the receiver is a tiny point in space whereas the transmitter has to send energy all over the space, wasting most of it. A very small portion of the transmitted energy reaches the receiver. "Most people would be surprised to know that the power of the signals received by a mobile phone is a hundred billionth of a billionth watt!" says Ashok Jhunjhunwala of IIT, Madras.

The amount of 'information' that can be sent through a channel depends on the ratio of the signal power to the noise power at the receiver's end, as shown by Claude Shannon. This creates the engineering challenge of building receivers that are able to detect a very weak signal and separate it from noise. By the way, by 'noise' engineers do not mean the audible noise of the bazaar, but random electrical fluctuations in the handset caused by: heat and internal circuitry, background electromagnetic radiation coming from the ionosphere or high tension wires, or even other people's cell phones near yours. Such electrical noise becomes an audible 'hiss' in your radio set, for example.

A brute force method to solve the problem of improving the signal to noise ratio is to make the transmitter as powerful as possible so that a sufficiently strong signal reaches the receiver. But there are limitations to the power pumped by the transmitter, especially when your cell phone itself is a transmitter. These limitations come from two sources: the power and longevity of the battery in a portable set (which should ideally weigh a few grams), and health hazards from the effect of powerful microwaves on the human brain.

It has been suspected that prolonged exposure to powerful microwaves could lead to brain tumours. Even though information in this area is patchy, everyone is aware of the risk. Hence portable handsets held close to the ear are mandated to be of low power – less than a watt. In most cases they actually have about half a watt of power. The sum result: the spreading nature of waves reduces the data rates possible on wireless compared with wired networks.

Incidentally, mobile wireless personal communications are not new. A demonstration of such communications took place in 1898, when Guglielmo Marconi, a flamboyant showman, gave a running commentary of a regatta on the Hudson river in New York while broadcasting from a tug. Another incident, which made wireless communication the talk of the town, was the capture in 1910 of a criminal on board a ship, when the captain of the ship received a secret wireless message. In the 1920s and 1930s experiments were conducted to use radio-telephony in the military, in police departments and fire brigades in the US.

The main issue, one that had to be tackled before commercial wireless communications became possible, was spectrum. To this day, spectrum allocation remains a major challenge.

Spectrum is the most precious societal resource, according to wireless engineers. It is a portion of the electromagnetic spectrum, or band of frequencies, reserved for a particular service. Regulatory agencies internationally and in individual countries allocate spectrum for various uses. For example, if you look at the frequency allocation in India by the wireless planning committee of the department of telecom (see box), you will see that different frequencies are allotted for different services like radio, TV, marine, defence, aeronautics, cell phones, pagers, radar, police, satellite up-linking and down-linking, and so on. The purpose of such allocation is to ensure that one service does not interfere with another.

The recent history of the wireless industry is full of jockeying by different service providers to get as big a chunk of frequencies for themselves as possible. Governments in the US and Europe have also looked at spectrum as a resource to be auctioned, and have earned large sums of money that way. In these countries there is hardly any licence fee for starting a service, but you have to buy the right to use a particular frequency exclusively for your service.

Interestingly, the first real advance in 'multiplying' the spectrum has been the multiple input multiple output (MIMO) technology, which uses multiple antennae at both ends of the link. The spectrum is multiplied by the number of antenna pairs! This idea, originally proposed by Arogyaswami Paulraj of the Information Systems Laboratory at Stanford University, is now a major frontier for enhancing wireless systems.

MEGAHERTZ, GIGAHERTZ AND ALL THAT

Electromagnetic waves, first predicted by British scientist James Clerk Maxwell, were later experimentally discovered by H.R. Hertz, to commemorate which, one electromagnetic wave vibration is called a hertz (Hz). A KiloHz is a thousand hertz, a MegaHz(MHz) is a million hertz and a GigaHz (GHz) is a billion hertz.

The part of the spectrum used for communications and broadcasting is known as the radio frequency (RF) spectrum; it extends from about 10 kHz to about 30 GHz. The International Telecom Union, an intergovernmental body, regulates the allocation of different bands for various end uses worldwide.

The following table illustrates the primary use to which different parts of the spectrum are allocated by the wireless planning committee of the Indian government:

Frequency	*Frequency band*	*Use*
500KHz-1.6 MHz	Medium Wave	Radio broadcast, All India Radio
2MHz-28 MHz	Short Wave (HF)	Overseas radio broadcast, defence, diplomatic, corporate, police aviation
30 MHz-300 MHz	Very High frequency (VHF)	TV, police, paging, FM radio, aeronautical and maritime coomunications, trunk telecommunications
300 MHz-3 GHz	Ultra high frequency	TV, defence, aeronautical, railways, cellular mobile, global positioning,

	(UHF)	WLL, radar
3 GHz-7 GHz	C-band and extended C-band	Microwave links (DoT), VSAT, INSAT uplink and down link, civil and defence radars
7 GHZ-8.5 GHz	X-band	Mobile base stations, remote sensing satellites
10 GHz-30	Ku-band	Intracity microwave, inter-satellite Communications, direct- to-home broadcasting,
20 GHz-30 GHz	Ka-band	Broadband satellite service

It is the scarcity of available spectrum that has led to all of the major technological developments in wireless over the last fifty years. The idea of cellular telephones emerged from this scarcity. Cellular telephony is actually very simple, and was articulated as far back as the 1940s by scientists at Bell Labs. Let me illustrate it with a simple example (the figures are illustrative, not realistic).

Every voice channel needs a certain bandwidth. Thus, within the allotted spectrum of, say, 10 MHz (850 MHz–860 MHz) only 5,000 calls can be handled at one time, assuming a highly compressed voice channel of only 2 kHz (16 kbit). If, on an average only ten per cent of subscribers use the telephone at any given time, we can conclude that the network can support 50,000 users. This may be fine for a police force or fire brigade but definitely not for a commercial service in a large city.

THE CELLULAR IDEA

Cellular technology solves this conundrum by dividing a large city into cells containing, say, a thousand subscribers each. It uses 850-852.5 MHz in one cell, 852.5-855 MHz each in the surrounding six cells, and 855- 857.5 Mhz in the next layer of cells. Using transmitters of the right power, we can ensure that the effect of the first set of frequencies do not reach farther than the cell containing the transmitter and its immediate neighbours, so that the next circle of cells using 855-857.5 MHz do not receive anything from the 850-852.5 MHz transmitters. This assured, we can use the 850-852.5 MHz frequency again in the fourth set of cells safely without any danger of interference.

By planning a sufficiently small and dense cellular structure, we can cover a million subscribers in a city like Mumbai using only 7.5 MHz of spectrum and keeping the remaining 2.5 MHz for emergencies or sudden surges in demand. Making use of the short range of microwaves, cellular architecture allows for the reuse use of the same set of frequencies, thereby increasing capacity.

If the transmitting power of the cell phone is so low, then how does the call reach somebody who is miles away and is moving? First of all each cell has a base station with which the caller communicates. This base station is connected to other base stations and finally to the mobile switching centre. When a mobile caller activates his handset, the base station recognises the subscriber through an automatically generated signal, checks the services he is eligible for and notifies the network that the caller is in this particular cell. Then it sends the message to the switch that he wishes to talk to another subscriber with a number.

The mobile switching centre talks to different base stations, finds out where the receiver is at that moment and connects the caller's base station with the base station in whose territory the receiver is available. The connection is made. Meanwhile, the caller might move out of the sphere of influence of the first base station and into that of its neighbouring base station. If that happens, the neighbouring station first detects his approach, assigns a new set of frequencies to him (remember, neighbouring base stations use different frequencies) and continues his call without interruption. This is called a hand-off. All this takes place in milliseconds, and neither the caller nor the receiver is aware that a transfer has taken place. When the call is over, the connection is broken and the information is sent to the records for billing purposes.

ROAM AWAY

When you travel to a different city or state which has a different service provider, a 'roaming' facility is provided. Under this the local cell phone company in that city acts as a conduit for the calls you are making; you get one single bill as if you were moving through one continuous, ubiquitous network. The accounting and sharing of revenues by the two companies is not visible to the customer.

Historically, several analog cellular services came into being in the US and Europe in the 1970s. In a continent like Europe where travelling a few hundred miles can take you through several countries, the question of a system with a smooth roaming facility became important. A group was set up in the early 1980s to study the issue and prescribe standards. The group, which came out with the new generation of digital cellular standards was called Groupe Speciale Mobile (GSM).

GSM is widely used in India, China, all of Asia (except Japan), Oceania and Europe. A characteristic of this technology is that all the subscriber information is in a smart chip called the SIM card, which is inserted in the handset. If you wish to change your handset, you simply remove the SIM card from the old handset, insert it in your new handset and you are all set to make or receive a call.

While Europe created a standard through consultations among experts and imposed it on everybody, the US took a different route. It has allowed the use of several different technologies, and expected the market to decide which is better. Thus analog (AMPS), PCS, GSM (with a different spectrum than in India) and CDMA coexist in the US. But since the technologies are widely different, it becomes expensive, and at times impossible, to roam all over the US unless your particular network also has its coverage where you currently are.

OPPORTUNITIES IN CHAOS

The US was the first country to deregulate telecom and introduce competition. The process started with the breaking up of AT&T in the early 1980s and has continued to this day. The introduction of competition and the restriction of monopolies have benefited customers in a big way. But this has, at times, led to piquant situations. Companies that understood the technology were not allowed to offer the service and new players were allowed entry even when their sole qualification was that they did not know the technology! Thus, when cellular licences were first released in the US, as an anti-monopoly measure, AT&T, which had pioneered cellular architecture, was not allowed to offer a cellular service. The smaller companies that were allowed were new entrepreneurs who did not have expertise in cellular technology.

This potentially chaotic situation spelt opportunities for some people. One smart Indian wireless engineer who exploited the opportunity was Rajendra Singh. "I had just finished my PhD in wireless technology and started teaching in Kansas; and my wife Neera, a chemical engineer, was doing her Master's there," recalls Singh at his beautiful mansion on the banks of the Potomac river in Washington, DC. At that time entrepreneurs who wanted to start a cellular service were supposed to submit their network plan to the Federal Communications Commission (FCC), the telecom regulatory body in the US. But independent experts to file and evaluate the bids were lacking.

"I started helping some of them," says Singh. "Since Neera knew computer programming, we developed software on a simple IBM PC to work out base station placement to get uniform coverage. We also developed simple equipment that could actually measure the signal at various points in the area and check the theoretical calculations. We sent the plan to a company which had won the licence for the Baltimore area near Washington DC. The company called back and asked how much we wanted to be paid for this. I said we did not want anything. It is just a piece of simple calculation that we did. But the guy said, 'No, you have to accept some money for this'. I said, 'OK, I shall charge $1,500'. The company wanted me to immediately shift to Washington and join them. I was told that other consultants had asked for six months' time and a fee of $80,000 to do what we had done overnight with our software."

It was not always so luxurious for Singh, who came from a backward village, Kairoo, in Rajasthan, which had neither electricity nor telephones. He lost an eye in a childhood accident due to lack of medical facilities in the village. Singh studied electrical engineering at IIT, Kanpur, went to the US in 1975 for his PhD, and proved to be a smart engineer who built a fortune using appropriate technology. This Indian engineering couple effectively became the architects of most US cellular networks in the 1980s. Later, in the 1990s, their consulting company, LCC, spread its wings to over forty countries.

While optimum use of spectrum became Mr & Mrs Singh's bread, butter and jam, there was another trend that violated all common sense in wireless engineering. It was called spread spectrum. Its champions said they would use the entire available spectrum to send a message. For wireless engineers weaned from childhood on 'communication channels', this was sacrilege. Interestingly, the champion of this technology was a Hollywood actress.

BEAUTY WITH BRAINS

Hedy Lamarr hit the headlines as an actress with a nude swimming scene in the Czech film *Ecstasy* (1933). She then married a rich pro-Nazi arms merchant, Fritz Mandl. For Mandl, she was a trophy wife, who he took along to parties and dinners to mingle with the high and mighty in Europe's political, military and business circles. But Hedy was no bimbo. Little did he suspect that beneath the beautiful exterior lay a sharp brain with an aptitude for technology! Lamarr was able to pick up quite a bit of the technical shoptalk of the men around the table.

When the Second World War began, Lamarr, a staunch anti-Nazi escaped to London. There she convinced Louis Mayer of MGM Studios to sign her up. Mayer, having heard of her reputation after *Ecstacy*, advised her to change her name from Hedwig Eva Marie Kiesler to Hedy Lamarr and to act in "wholesome family movies", which she promptly agreed to.

As the war progressed and the US joined the UK and the Soviet Union after the Japanese attack on Pearl Harbour, Lamarr informed the US government that she was privy to a considerable amount of Axis war technology and she wanted to help. The defence department had little faith in her claims and advised her, instead, to sell war bonds to rich Americans. But Lamarr was unrelenting. Along with her friend George Antheil, an avant-garde composer and musician, she patented their 'secret communication system' and gave the patent rights free to the US military. The patent was about a design for a jamming-free radio guidance system for submarine-launched torpedoes based on the frequency-hopping spread-spectrum technique.

Lamarr's idea consisted of two identical punched paper rolls. One roll was located in the submarine, and changed the transmission frequency as it was rotated. The other, embedded in the torpedo, also rotated and hopped to the appropriate receiving frequency. The enemy jammer would thus be left guessing about the guiding frequency. The idea, which came to be named frequency hopping, was ingenious but the US navy was not technologically advanced enough to use it!

SPREADING OUT WITH SPREAD SPECTRUM

In the late 1950s, as digital computers appeared on the scene, the US Navy revived its interest in Lamarr's ideas. With the development of microchips and digital communications, advanced and secure communications systems have been developed for military purposes using spread spectrum techniques. Since this technology can be used for secure communications, which cannot be jammed or intercepted, the US military has done extensive research and development in it since the 1960s.

In the telecom revolution of the 1990s, these techniques have been used to develop civilian applications in cellular phones, wireless in local loop, personal communication systems, and so on. The unlikely inventor showed that if you have a sharp brain, party hopping can lead to frequency hopping!

Spread spectrum technology assures a high level of security and privacy in wireless communication. It came into wide usage in the 1990s as Qualcomm demonstrated its successful application for cellular phones. Another anti-snooping technique involves signals being mixed with strong doses of 'noise', and then transmitted. Only the intended receiver knows the exact characteristics of the 'noise' that has been added, and can subtract it from the received signal, thereby recovering the transmitted signal. This technique works best when the added 'noise' is very powerful.

Qualcomm used this technique to develop its CDMA technology, which is not only inherently secure but also less prone to a common, 'multipath' problem, or fading in and fading out of voice in cell phones. The problem occurs because of the signal getting reflected by natural and man-made structures and reaching the receiver at different times, causing interference and the fading-in and fading-out effect. However, multipath is a frequency-dependent effect; hence it does not affect spread spectrum based systems as the broadcast is made not at one frequency, but a whole bunch of them in a wide band.

In the late 1990s, before their break-up, Hollywood stars Tom Cruise and Nicole Kidman were deeply upset when a man used a commonly available frequency scanner to find out what frequency their cellular phones were using. He then proceeded to snoop into their private conversations, tape them and sell them to an American tabloid. The episode brought to light the lack of privacy in an analog cellular phone call, and stressed the advantage of cell phones using digital spread spectrum technology like CDMA.

Because of their high costs and tariffs, cellular phones were initially popular only among the rich and powerful. In the early 1990s you could see cell phones mainly with a small set of people: senior executives, stock brokers, politicians, film stars, etc. But, as costs drop, they are finding increasing use among ordinary people everywhere.

In a country like India cellular phones are not a luxury, but a necessity for a large section of the middle class and lower middle class population, including the self-employed, be they taxi drivers, carpenters, plumbers, electricians, roadside mechanics, salesmen, medical representatives or couriers. Either their profession makes them continuously mobile or they do not have shops and offices. Even if they do have an office, you might ask, what is the point in having a fixed phone there, when they are out servicing customers? That is why, for many Indians, the telecom revolution translates to STD booths and affordable cell phones.

FURTHER READING

1. Telecommunications and the Computer—James Martin, third edition, Prentice Hall of India Pvt Ltd, 1992.
2. Mobile Communications—Jochen Schiller, Addison-Wesley, 2000.
3. Digital Telephony—John C Bellamy, third edition, John Wiley and Sons Inc. 2000.
4. The Telecom Story and the Internet—Mohan Sundara Rajan, fourth edition, National Book Trust, India, 2001.
5. Development Democracy, and the Village Telephone—Sam Pitroda, *Harvard Business Review*, Nov-Dec 1993.
6. Vision Values and Velocity—Sam Pitroda, *Siliconindia,* 2001.
7. Desktop Encyclopedia of Telecommunications—Nathan J Muller, second edition, McGraw-Hill, 2000.
8. A Mathematical Theory of Communication—Claude E. Shannon, *Bell Labs Technical Journal,* volume 27, pp 379-423, 623-656, July, October 1948 (www.cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf).
9. Physics and the communications industry—Bill Brinkman and Dave Lang (http://www.bell-labs.com/about/history/physicscomm/).
10. An overview of Information Theory Bell Labs, 1998 (http:// www.lucent.com/minds/infotheory/what5.html).
11. "Extra Terrestrial Relays: Can Rocket Stations Give World-wide Radio Coverage"—Arthur C Clarke, *Wireless World,* October 1945, pp 305-308. (http://www.sciencemuseum.org.uk/on-line/clarke/ww2.asp).
12. "Space: The New Business Frontier"—Shivanand Kanavi, *Business India,* April 21- May 4, 1997. (http://reflections-shivanand.blogspot.com/2007_09_28_archive.html)
13. "The Indians Want In"—Shivanand Kanavi, *Business India,* Nov 16-29, 1998.
14. "Those Magnificent Men…"—Shivanand Kanavi, *Business India,* Nov 21- Dec 4, 1994. (http://reflections-shivanand.blogspot.com/2007_10_03_archive.html)
15. "Remotely sensing profits"—Shivanand Kanavi, *Business India,* Feb 28- March 13, 1994. (http://reflections-shivanand.blogspot.com/2007/09/isro-remote-sensing.html)
16. "Reaching out with spread spectrum"—Shivanand Kanavi, *Business India,* Jan 25-Feb 7, 1999. (http://reflections-shivanand.blogspot.com/2007/08/hollywood-beauty-and-cdma.html)
17. "Selling vacuum"—Shivanand Kanavi, *Business India,* Dec 14-27, 1998 18. "Interconnect and prosper"—Shivanand Kanavi, *Business India,* Sep 20- Oct 3, 1999.

Optical technology: Lighting up our lives

"Behold, that which has removed the extreme in the pervading darkness, Light became the throne of light, light coupled with light "

—**Basavanna**, twelfth century, Bhakti poet, Karnataka

"A splendid light has dawned on me about the absorption and emission of radiation."

—**Albert Einstein**, in a letter to Michael Angelo Besso,
in November 1916

"Roti, kapda, makaan, bijlee aur bandwidth (food, clothing, housing, electricity and bandwidth) will be the slogan for the masses."

—**Dewang Mehta**, an IT evangelist

It is common knowledge that microchips are a key ingredient of modern IT. But optical technology, consisting of lasers and fibre optics, is not given its due. This technology already affects our lives in many ways; it is a vital part of several IT appliances and a key element in the modern communication infrastructure.

Let us look at lasers first. In popular perception, lasers are still identified with their destructive potential—the apocalyptic 'third eye' of Shiva. It was no coincidence that the most powerful neodymium glass laser built for developing nuclear weapon technology at the Lawrence Livermore Laboratory in the US (back in 1978) was named Shiva.

Villains used lasers in the 1960s to cut the vaults of Fort Knox in the James Bond movie, *Goldfinger.* In 1977 Luke Skywalker and Darth Vader had their deadly duels with laser swords in the first episode of *Star Wars.* As if to prove that life imitates art, Ronald Reagan poured millions of dollars in the 1980s, in a 'ray gun', a la comic-strip super-hero stories, with the aim of building the capability to shoot down Soviet nuclear missiles and satellites.

THE BENIGN THIRD EYE

In our real, daily lives, lasers have crept in without much fanfare:

- All sorts of consumer appliances, including audio and video CD players and DVD players use lasers.

- Multimedia PCs are equipped with a CD-ROM drive which uses a laser device to read or write.

- Light emitting diodes (LED)—country cousins of semiconductor lasers—light up digital displays and help connect office computers into local area networks.

- LED-powered pointers have become popular in their use with audiovisual presentations.

- Who can forget the laser printer that has revolutionised publishing and brought desk top publishing to small towns in India?
- Laser range finders and auto-focus in ordinary cameras have made 'expert photographers' of us all.

- The ubiquitous TV remote control is a product of infrared light emitting diodes.

- The bar code reader, used by millions of sales clerks and storekeepers, and in banks and post offices, is one of the earliest applications of lasers.

- Almost all overseas telephone calls and a large number of domestic calls whiz through glass fibres at the speed of light, thanks to laser-powered communications.

- The Internet backbone, carrying terabits (tera = 10^{12}, a million million) of data, uses laser-driven optical networks.

C.K.N. Patel won the prestigious National Medal of Science in the US in 1996 for his invention of the carbon dioxide laser, the first laser with high-power applications, way back in 1964 at Bell Labs. He says, "Modern automobiles have thousands of welds, which are made by robots wielding lasers. Laser welds make the automobile safer and lighter by almost a quintal. Even fabrics in textile mills and garment factories are now cut with lasers."

Narinder Singh Kapany, the inventor of fibre optics, was also the first to introduce lasers for eye surgery. He did this in the 1960s along with doctors at Stanford University. Today's eye surgeons are armed with excimer laser scalpels that can make incisions less than a micron (thousandth of a millimetre) wide, on the delicate tissues of the cornea and retina.

LASER BASICS

So what are lasers, really? They produce light that has only one wavelength and a high directionality and is coherent. What do these things mean, and why are they important? Any source of light, man-made or natural, gives out radiation that is a mixture of wavelengths—be it a kerosene lantern, a wax candle, an electric bulb or the sun. Different wavelengths of light correspond to different colours.

When atoms fly around in a gas or vibrate in a solid in random directions, the light (photons) emitted by them does not have any preferred direction; the photons fly off in a wide angle. We try to overcome the lack of direction by using a reflector that can narrow the beam, as from torchlight, to get a strong directional beam. However, the best of searchlights used, say, outside a circus tent or during an air raid, get diffused at a distance of a couple of miles.

The intensity of a spreading source at a distance of a metre is a hundred times weaker than that at ten centimetres and a hundred million times weaker at a distance of one kilometre. Since there are physical limits to increasing the strength of the source, not to mention the prohibitive cost, we need a highly directional beam. Directionality becomes imperative for long-distance communications over thousands of kilometres..

Why do we need a single wavelength? When we are looking for artificial lighting, we don't. We use fluorescent lamps (tube lights), brightly coloured neon signs in advertisements or street lamps filled with mercury or sodium. All of them produce a wide spectrum of light. But a single wavelength source literally provides a vehicle for communications. This is not very different from the commuter trains we use en masse for efficient and high-speed transportation. Understandably, telecom engineers call them 'carrier waves'. In the case of radio or TV transmission, we use electronic circuits that oscillate at a fixed frequency. Audio or video signals are superimposed on these carrier channels, which then get a commuter ride to the consumer's receiving set.

With electromagnetic communications, the higher the frequency of the carrier wave, the greater the amount of information that can be sent piggyback on it. Since the frequency of light is million times greater than that of microwaves, why not use it as a vehicle to carry our communications? It was this question that led to optical communications, where lasers provide the sources of carrier waves, electronics enables your telephone call or Internet data to ride piggyback on it, and thinner-than hair glass fibres transport the signal underground and under oceans.

We have to find ways to discipline an unruly crowd of excited atoms and persuade them to emit their photons in some order so that we obtain monochromatic, directional and coherent radiation. Lasers are able to do just that.

Why coherence? When we say a person is coherent in his expression, we mean that the different parts of his communication, oral or written, are connected logically, and hence make sense. Randomness, on the other hand, cannot communicate anything. It produces gibberish.

If we wish to use radiation for communications, we cannot do without coherence. In radio and other communications this was not a problem since the oscillator produced coherent radiation. But making billions of atoms radiate in phase necessarily requires building a new kind of source. That is precisely what lasers are.

Like many other ideas in modern physics, lasers germinated from a paper by Albert Einstein in 1917. He postulated that matter could absorb energy in discrete quanta if the size of the quantum is equal to the difference between a lower energy level and a higher energy level. The excited atoms, he noted, can come down to the lower energy state by emitting a photon of light spontaneously.

On purely theoretical considerations, Einstein made a creative leap by contending that the presence of radiation creates an alternative way of de-excitation, called stimulated emission. In the presence of a photon of the right frequency, an excited atom is induced to emit a photon of the exact same characteristics. Such a phenomenon had not yet been seen in nature.

Stimulated emission is like the herd effect. For example, a student may be in two minds about skipping a boring lecture, but if he bumps into a couple of friends who are also cutting classes, then he is more likely to join the gang.

A most pleasant outcome of this herd behaviour is that the emitted photon has the same wavelength, direction and phase as the incident photon. Now these two photons can gather another one if they encounter an excited atom. We can thus have a whole bunch of photons with the same wavelength, direction, and phase. There is one problem, though; de-excited atoms may absorb the emitted photon, and hence there may not be enough coherent photons coming out of the system.

What if the coherent photons are made to hang around excited atoms long enough without exiting the system in a hurry? That will lead to the same photon stimulating more excited atoms. But how do you make photons hang around? You cannot slow them down. Unlike material particles like electrons, which can be slowed down or brought to rest, photons will always zip around with the same velocity (of light of course!)—300,000 km per second.

THE BARBERSHOP SOLUTION

Remember the barber’s shop, with mirrors on opposite walls showing you a large number of reflections? Theoretically, you could have an infinite number of reflections, as if light had been trapped between parallel facing mirrors. Similarly, if we place two highly polished mirrors at the two ends of our atomic oscillator, coherent photons will reflected back and forth, and we will get a sustainable laser action despite the usual absorptive processes.

At the atomic level, of course, we need to go further than the barber’s shop. We need to adjust the mirrors minutely so that we can achieve resonance, i.e., when the incident and reflected photons match one another in phase, and standing waves are formed. Lo and behold, we have created a light amplification by stimulated emission of radiation (laser).

In the midst of disorderly behaviour we can see order being created by a laser. Physics discovered that the universe decays spontaneously into greater and greater disorder. If you are a stickler, 'the entropy—measure of disorder—of an isolated system can only increase'. This is the second law of thermodynamics. So are we violating this law? Are we finally breaking out of thermodynamic tyranny?

It should be noted, however, that the universe becomes interesting due to the creation of order. Evolution of life and its continuous reproduction is one of the greatest acts of creating order. However, rigorous analysis shows that even when order is created in one part of the universe, on the whole, disorder increases. Lasers are humanity's invention of an order-creating system.

Charles Townes, a consultant at Bell Labs, first created microwave amplification through stimulated emission in 1953. He called the apparatus a maser. Later work by Townes and Arthur Schawlow at Bell Labs, and Nikolay Basov and Aleksandr Prokharov in the Soviet Union led to the further development of laser physics. Townes, Basov and Prokharov were awarded the Nobel Prize for their work in 1964. Meanwhile, in 1960, Theodore Maiman, working at the Hughes Research Laboratory, had produced the first such instrument for visible light—hence the first laser—using a ruby crystal.

Since then many lasing systems have been created. At Bell Labs C K N Patel did outstanding work in gas lasers and developed the carbon dioxide laser in 1964. This was the first high power continuous laser and since then it has been perfected for high power applications in manufacturing.

THE SEMICONDUCTOR REVOLUTION IN LASERS

What made lasers become hi-tech mass products was the invention of semiconductor lasers in 1962 by researchers at General Electric, IBM, and the MIT Lincoln Laboratory. These researchers found that diode devices based on the semiconductor gallium arsenide convert electrical energy into light. They were highly efficient in their amplification, miniature in size and eventually inexpensive. These characteristics led to their immediate application in communications, data storage and other fields.

Today, the performance of semiconductor lasers has been greatly enhanced by using sandwiches of different semiconductor materials. Such 'hetero-junction' lasers can operate even at room temperature, whereas the older semiconductor lasers needed cooling by liquid nitrogen (to around -77 ^{0}C). Herbert Kroemer and Zhores Alferov were awarded the Nobel Prize in physics in 2000 for their pioneering work in hetero-structures in semiconductors. Today, various alloys of gallium, arsenic, indium, phosphorus and aluminium are used to obtain the best LEDs and lasers.

One of the hottest areas in semiconductor lasers is quantum well lasers, or cascade lasers. This area came into prominence with the development of techniques of growing semiconductors layer by layer using molecular beam epitaxy. Researchers use this technique to work like atomic bricklayers. They build a laser by placing a layer of a semiconductor with a particular structure and then placing another on top with a little bit of cementing material in between. By accurately controlling the thickness of these layers and their composition, researchers can adjust the band gaps in different areas. This technique is known as 'band gap engineering'.

If the sandwich is thin enough, it acts as a quantum well for electrons. The electrons confined in this way lead to quantum systems called quantum wells (also known as particle in a box). The gap in the energy levels in such quantum wells can be controlled minutely and used for constructing a laser. Further, by constructing a massive club sandwich, as it were, we can have several quantum wells next to each other. The electron can make a stimulated emission of a photon by jumping to a lower level in the neighbouring well and then the next one and so on. This leads to a cascade effect like a marble dropping down a staircase. The system ends up emitting several photons of different wavelengths, corresponding to the quantum energy staircase. Frederico Capasso and his team built the first such quantum cascade laser at Bell Labs in 1994.

Once a device can be made from semiconductors, it becomes possible to miniaturise them while raising performance levels and reducing their price. That's the pathway to mass production and use. This has happened in the case of lasers too.

We can leave the physics of lasers at this point and see how lasers are used in appliances of daily use: A bar-code reader uses a tiny helium-neon laser to scan the code. A detector built into the reader detects reflected light and the white-and black bars are then converted to a digital code that identifies the object.

A laser printer uses static electricity; that's what makes your polyester shirt or acrylic sweater crackle sometimes. The drum assembly inside the laser printer is made of material that conducts when exposed to light. Initially, the rotating drum is given a positive charge. A tiny movable mirror reflects a laser beam on to the drum surface, thereby rendering certain points on the drum electrically neutral. A chip controls the movement of the mirror. The laser 'draws' the letters and images to be printed as an electrostatic image.

After the image is set, the drum is coated with positively charged toner (a fine, black powder). Since it has a positive charge, the toner clings to the discharged areas of the drum, but not to the positively charged 'background'. The drum, with this powder pattern, rolls over a moving sheet of paper that has already been given a negative charge stronger than the negative charge of the image. The paper attracts the toner powder. Since it is moving at the same speed as the drum, the paper picks up the image exactly. To keep the paper from clinging to the drum, it is electrically discharged after picking up the toner. Finally, the printer passes the paper through a pair of heated rollers. As the paper passes through these rollers, the toner powder melts, fusing with the paper, which is why pages are always warm when they emerge from a laser printer.

Compact discs are modern *avatars* of the old vinyl long-playing records. Sound would be imprinted on the LPs by a needle as pits and bumps. When the needle in the turntable head went over the track, it moved in consonance with these indentations. The resultant vibrations were amplified mechanically to reproduce the sound we heard as music. Modern-day CDs and DVDs are digital versions of the old Edison's phonograph. Sound or data is digitised and encoded in tiny black or white spots corresponding to ones and zeros. These spots are then embedded in tiny bumps that are 0.5 microns wide, 0.83 microns long and 0.125 micron high. The bumps are laid out in a spiral track much as in the vinyl record. A laser operating at a 0.780-micron wavelength lights up these spots and the reflected signal is then read by a detector as a series of ones and zeroes, which are translated into sound.

In the case of DVDs, or digital versatile discs, the laser operates at an even smaller wavelength, and is able to read much smaller bumps. This allows us to increase the density of these bumps in the track on a DVD with more advanced compression and coding techniques. This means we can store much more information on a DVD than we can on a CD. A DVD can store several GB of information compared with the 800 MB of data a CD can store.

A CD is made from a substratum of polycarbonate imprinted with microscopic pits and coated with aluminium, which is then protected by a thin layer of acrylic. The incredibly small dimensions of the bumps make the spiral track on a CD almost five kilometres long! On DVDs, the track is almost twelve kilometres long.

To read something this small you need an incredibly precise discreading mechanism. The laser reader in the CD or DVD player, which has to find and read the data stored as bumps, is an exceptionally precise device.

The fundamental job of the player is to focus the laser on the track of bumps. The laser beam passes through the polycarbonate layer, reflects off the aluminium layer, and hits an opto-electronic device that detects changes in light. The bumps reflect light differently than the rest of the aluminium layer, and the opto-electronic detector senses the change in reflectivity. The electronics in the drive interpret the changes in reflectivity in order to read the bits that make up the bytes. These are then processed as audio or video signals.

With the turntables of yesterday's audio technology, the vibrating needles would suffer wear and tear. Lasers neither wear themselves out nor scratch the CDs, and they are a thousand times smaller than the thinnest needle. That is the secret of high-quality reproduction and the high quantity of content that can be compressed into an optical disc.

C.K.N. Patel recalls how, in the 1960s, the US defence department was the organisation that evinced the greatest interest in his carbon dioxide laser. "The launch of the Sputnik by the Soviet Union created virtual panic," he says. "That was excellent, since any R&D project which the military thought remotely applicable to defence got generously funded." 'Peacenik' Patel, who is passionate about nuclear disarmament, is happy to see that the apocalyptic 'Third Eye' has found peaceful applications in manufacturing and IT. Patel refuses to retire and is busy, in southern California, trying to find more applications of lasers for health and pollution problems.

To get into the extremely important application of lasers in communications, we need to look at fibre optics more closely.

DUG UP ROADS

Outside telecom circles, fibre optics is not very popular among city dwellers in India. Because, in the past couple of years, hundreds of towns and cities in India have been dug up on an precedented scale. The common refrain is: "They are laying fibre-optic cable". Fibre optics has created an obstacle course for pedestrians and drivers while providing grist to the mills of cartoonists like R.K. Laxman. Being an optimist, I tell my neighbours, "Soon we will have a bandwidth infrastructure fit for the twenty-first century." What is bandwidth? It is an indication of the amount of information you can receive per second, where 'information' can mean words, numbers, pictures, sounds or films.

Bandwidth has nothing to do with the diameter of the cable that brings information into our homes. In fact, the thinnest fibres made of glass— thinner than human hair—can bring a large amount of information into our homes and offices at a reasonable cost. And that is why fibre optics is playing a major role in the IT revolution.

It is only poetic justice that words like fibre optics are becoming popular in India. Very few Indians know that an Indian, Narinder Singh Kapany, a pioneer in the field, coined them in 1960. We will come to his story later on, but before that let us look at what fibre optics is. It all started with queries like: Can we channel light through a curved path, even though we know that light travels in a straight line? Why is that important? Well, suppose you want to examine an internal organ of the human body for diagnostic or surgical purposes. You would need a flexible pipe carrying light. Similarly, if you want to communicate by using light signals, you cannot send light through the air for long distances; you need a flexible cable carrying light over such distances.

The periscopes we made as class projects when we were in school, using cardboard tubes and pieces of mirror, are actually devices to bend light. Bending light at right angles as in a periscope was simple. Bending light along a smooth curve is not so easy. But it can be done, and that is what is done in optic fibre cables.

For centuries people have built canals or viaducts to direct water for irrigation or domestic use. These channels achieve maximum effect if the walls or embankments do not leak. Similarly, if we have a pipe whose insides are coated with a reflecting material, then photons or waves can be directed along easily without getting absorbed by the wall material. A light wave gets reflected millions of times inside such a pipe (the number depending on the length and diameter of the pipe and the narrowness of the light beam). This creates the biggest problem for pipes carrying light. Even if we can get coatings with 99.99 per cent reflectivity, the tiny leakage’ of 0.01 per cent on each reflection can result in a near-zero signal after 10,000 reflections.

Here a phenomenon called total internal reflection comes to the rescue. If we send a light beam from water into air, it behaves peculiarly as we increase the angle between the incident ray and the perpendicular. We reach a point when any increase in the angle of incidence results in the light not leaving the water and, instead, getting reflected back entirely. This phenomenon is called total internal reflection. Any surface, however finely polished, absorbs some light, and hence repeated reflections weaken a beam. But total internal reflection is a hundred per cent, which means that if we make a piece of glass as non-absorbent as possible, and if we use total internal reflection, we can carry a beam of light over long distances inside a strand of glass. This is the principle used in fibre optics.

The idea is not new. In the 1840s, Swiss physicist Daniel Collandon and French physicist Jacques Babinet showed that light could be guided along jets of water. British physicist John Tyndall popularised the idea further through his public demonstrations in 1854, guiding light in a jet of water flowing from a tank. Since then this method has been commonly used in water fountains. If we keep sources of light that change their colour periodically at the fountainhead, it appears as if differently coloured water is springing out of the fountain.

Later many scientists conceived of bent quartz rods carrying light, and even patented some of these inventions. But it took a long time for these ideas to be converted into commercially viable products. One of the main hurdles was the considerable absorption of light inside glass rods.

Narinder Singh Kapany recounted to the author, “When I was a high school student at Dehradun in the beautiful foothills of the Himalayas, it occurred to me that light need not travel in a straight line, that it could be bent. I carried the idea to college. Actually it was not an idea but the statement of a problem. When I worked in the ordnance factory in Dehradun after my graduation, I tried using right-angled prisms to bend light. However, when I went to London to study at the Imperial College and started working on my thesis, my advisor, Dr Hopkins, suggested that I try glass cylinders instead of prisms. So I thought of a bundle of thin glass fibres, which could be bent easily. Initially my primary interest was to use them in medical instruments for looking inside the human body. The broad potential of optic fibres did not dawn on me till 1955. It was then that I coined the term fibre optics.”

Kapany and others were trying to use a glass fibre as a light pipe or, technically speaking, a ‘dielectric wave guide’. But drawing a fibre of optical quality, free from impurities, was not an easy job.

Kapany went to the Pilkington Glass Company, which manufactured glass fibre for non-optical purposes. For the company, the optical quality of the glass was not important. “I took some optical glass and requested them to draw fibre from that,” says Kapany. “I also told them that I was going to use it to transmit light. They were perplexed, but humoured me.” A few months later Pilkington sent spools of fibre made of green glass, which is used to make beer bottles. “They had ignored the optical glass I had given them. I spent months making bundles of fibre from what they had supplied and trying to transmit light through them, but no light came out. That was because it was not optical glass. So I had to cut the bundle to short lengths and then use a bright carbon arc source."

Kapany was confronted with another problem. A naked glass fibre did not guide the light well. Due to surface defects, more light was leaking out than he had expected. To transmit a large image he would have needed a bundle of fibres containing several hundred strands; but contact between adjacent fibres led to loss of image resolution. Several people then suggested the idea of cladding the fibre. Cladding, when made of glass of a lower refractive index than the core, reduced leakages and also prevented damage to the core. Finally, Kapany was successful; he and Hopkins published the results in 1954 in the British journal *Nature*.

Kapany then migrated to the US and worked further in fibre optics while teaching at Rochester and the Illinois Institute of Technology. In 1960, with the invention of lasers, a new chapter opened in applied physics. From 1955 to 1965 Kapany was the lead author of dozens of technical and popular papers on the subject. His writings spread the gospel of fibre optics, casting him as a pioneer in the field. His popular article on fibre optics in the *Scientific American* in 1960 finally established the new term (fibre optics); the article constitutes a reference point for the subject even today. In November 1999, *Fortune* magazine published profiles of seven people who have greatly influenced life in the twentieth century but are unsung heroes. Kapany was one of them.

BELL WAS THERE, TOO

If we go back into the history of modern communications involving electrical impulses, we find that Alexander Graham Bell patented an optical telephone system in 1880. He called this a 'photophone'. Bell converted speech into electrical impulses, which he converted into light flashes. A photosensitive receiver converted the signals back into electrical impulses, which were then converted into speech. But the atmosphere does not transmit light as reliably as wires do; there is heavy atmospheric absorption, which can get worse with fog, rain and other impediments. As there were no strong and directional light sources like lasers at that time, optical communications went into hibernation. Bell's earlier invention, the telephone, proved far more practical. If Bell yearned to send signals through the air, far ahead of his time, we cannot blame him; after all, it's such a pain digging and laying cables.

In the 1950s, as telephone networks spread, telecommunications engineers sought more transmission bandwidth. Light, as a carrying medium, promised the maximum bandwidth. Naturally, optic fibres attracted attention. But the loss of intensity of the signal was as high as a decibel per metre. This was fine for looking inside the body, but communications operated over much longer distances and could not tolerate losses of more than ten to twenty decibels per kilometre.

Now what do decibels have to do with it? Why is signal loss per kilometre measured in decibels? The human ear is sensitive to sound on a logarithmic scale; that is why the decibel scale came into being in audio engineering, in the first place. If a signal gets reduced to half its strength over one kilometre because of absorption, after two kilometres it will become a fourth of its original strength. That is why communication engineers use the decibel scale to describe signal attenuation in cables.

In the early 1969s signal loss in glass fibre was one decibel per metre, which meant that after traversing ten metres of the fibre the signal was reduced to a tenth of its original strength. After twenty metres the signal was a mere hundredth its original strength. As you can imagine, after traversing a kilometre no perceptible signal was left.

A small team at the Standard Telecommunications Laboratories in the UK was not put off by this drawback. This group was headed by Antoni Karbowiak, and later by a young Shanghai-born engineer, Charles Kao. Kao studied the problem carefully and worked out a proposal for long-distance communications through glass fibres. He presented a paper at a London meeting of the Institution of Electrical Engineers in 1966, pointing out that the optic fibre of those days had an information-carrying capacity of one GHz, or an equivalent of 200 TV channels, or more than 200,000 telephone channels. Although the best available low-loss material then showed a loss of about 1,000 decibels/kilometre (dB/km), he claimed that materials with losses of just 10-20 dB/km would eventually be developed.

With Kao almost evangelistically promoting the prospects of fibre communications, and the British Post Office (the forerunner to BT) showing interest in developing such a network, laboratories around the world tried to make low-loss fibre. It took four years to reach Kao's goal of 20 dB/km. At the Corning Glass Works (now Corning Inc.), Robert Maurer, Donald Keck and Peter Schultz used fused silica to achieve the feat. The Corning breakthrough opened the door to fibre-optic communications. In the same year, Bell Labs and a team at the Ioffe Physical Institute in Leningrad (now St Petersburg) made the first semiconductor lasers, able to emit a continuous wave at room temperature. Over the next several years, fibre losses dropped dramatically, aided by improved fabrication methods and by the shift to longer wavelengths where fibres have inherently lower attenuation. Today's fibres are so transparent that if the Pacific Ocean, which is several kilometres deep, were to be made of this glass we could see the ocean bed!

Note one point here. The absorption of light in glass depends not only on the chemical composition of the glass but also on the wavelength of light that is transmitted through it. It has been found that there are three windows with very low attenuation: one is around 900 nanometres, the next at 1,300 nm and the last one at 1,550 nm. Once engineers could develop lasers with those wavelengths, they were in business. This happened in the 1970s and 1980s, thanks to Herbert Kroemer's hetero-structures and many hard-working experimentalists.

REPEATERS AND 'CHINESE WHISPERS'

All telephone systems need repeater stations at every few kilometres to receive the signal, amplify it and re-send it. Fibre optic systems need stations every few kilometres to receive a weak light signal, convert it into electronic signal, amplify it, use it to modulate a laser beam again, and re-send it. This process is exposed to risk of noise and errors creeping into the signal; the system needs to get rid of the noise and re-send a fresh signal. It is like a marathon run, where the organisers place tables with refreshing drinks all along the route so that the tired and dehydrated runners can refresh themselves. This means a certain delay, but the refreshment is absolutely essential.

Submarine cables must have as few points as possible where the system can break down because, once the cable is laid several kilometres under the sea, it becomes virtually impossible to physically inspect faults and repair them.

The development, in the 1980s, of fibre amplifiers, or fibres that act as amplifiers, has greatly facilitated the laying of submarine optic fibre cables. This magic is achieved through an innovation called the erbium doped fibre amplifier. Sections of fibre carefully doped with the right amount of erbium—a rare earth element—act as laser amplifiers.

While fibre amplifiers reduce the requirement of repeater stations, they cannot eliminate the need for them. That is because repeater stations not only amplify the signal, they also clean up the noise (whereas fibre amplifiers amplify the signal, noise and all). In fact, they add a little bit of their own noise. This is like the popular party game called Chinese whispers. If there is no correction in between, the message gets transmitted across a distance, but in a highly distorted fashion.

Can we get rid of these repeater stations altogether and send a signal which does not need much amplification or error correction over thousands of kilometres? That's a dream for every submarine cable company, though perhaps not a very distant one.

The phenomenon being used in various laboratories around the world to create such a super-long-distance runner is called a 'soliton' or a solitary wave. A Dutch gentleman first observed solitary waves nearly 300 years ago while riding along the famous canals of the Netherlands. He found that as boats created waves in canals, some waves were able to travel enormously long distances without dissipating themselves. They were named solitary waves, for obvious reasons. Scientists are now working on creating solitons of light that can travel thousands of kilometres inside optical fibres without getting dissipated.

As and when they achieve it, they will bring new efficiencies to fibre optic communications. Today, any signal is a set of waves differing in wavelength by very small amounts. Since the speeds of different wavelengths of light differ inside glass fibres, over a large distance the narrow packet tends to loosen up, with some portion of information appearing earlier and some later. This is called 'dispersion', something similar to the appearance of a colourful spectrum when light passes through a glass prism or a drop of rain. Solitons seem to be unaffected by dispersion. Long-distance cable companies are eagerly awaiting the conversion of these cutting-edge technologies from laboratory curiosities to commercial propositions.

Coming down to earth, we find that even though fibre optic cable prices have crashed in recent years, the cost of terminal equipment remains high. That is why it is not yet feasible to lay fibre optic cable to every home and office. For the time being, we have to remain content with such cables being terminated at hubs supporting large clusters of users, and other technologies being used to connect up the 'last mile' between the fibre optic network and our homes and offices.

FURTHER READING

1. "Zur Quantentheorie der strahlung" (Towards a quantum theory of radiation)—Albert Einstein, *Physikalische Zeitschrift*, Volume 18 (1917), pp 121-128 translated as "Quantum Theory of Radiation and Atomic Processes," in Henry A. Boorse and Lloyd Motz (eds.) The World of the Atom, Volume II, Basic Books, 1966, pp 884-901
2. Charles Townes—Nobel Lecture, 1964 (www.nobel.se/physics/laureates/ 1964/townes-lecture.html)
3. N.G. Basov—Nobel Lecture, 1964 (http://www.nobel.se/physics/laureates/1964/basov-lecture.html)
4. Lasers Theory and Applications—K. Thyagarajan, A.K. Ghatak, Macmillan India Ltd, 2001.
5. Chaos, Fractals and Self-Organisation: New perspectives on complexity in nature—Arvind Kumar, National Book Trust, India, 1996
6. Semiconductor devices Basic principle—Jasprit Singh, John Wiley, 2001.
7. Herbert Kroemer—Nobel Lecture, 2000 (www.nobel.se/physics/laureates/2000/kroemer-lecture.html).
8. Zhores Alferov—Nobel Lecture, 2000 (www.nobel.se/physics/laureates/2000/alferov-lecture.html).
9. Diminishing Dimensions—Elizabeth Corcoran and Glenn Zorpette,*Scientific American,* Jan 22, 1998.
10. Fiber Optics—Jeff Hecht, Oxford University Press, New York, 1999.
11. Fiber Optics—N.S. Kapany, *Scientific American,* November 1960.
12. Fibre Optics—G.K. Bhide, National Book Trust, India, 2000.
13. "Beyond valuations"—Shivanand Kanavi, *Business India,* Sep 17-30, 2001. (http://reflections-shivanand.blogspot.com/2007/08/tech-pioneers.html)

INTERNET

"Great Cloud. Please help me. I am away from my beloved and miss her very much. Please go to the city called Alaka where my beloved lives in our moonlit house"

—FROM *MEGHADOOT* (messenger cloud) of Kalidasa,
Sanskrit poet, playwright, fourth century AD

I am sure the Internet is on the verge of taking off in India to the next level of usage. I am not making this prediction based on learned market research by Gartner, Forrester, IDC or some other agency, but on observing my wife.

While I was vainly trying to get her interested in the PC and the Internet, I downloaded pages and pages of information on subjects of her interest. She thanked me but refused to get over her computer phobia. Not that she is anti-technology or any such thing. (In fact, she took to cell phones and SMS faster than me and showed me all kinds of tricks on her cell phone.) But, whenever I managed to bring her to the PC and turned on the Internet, she would say, "Who can stand this 'World Wide Wait'?!", and I would give up.

But a sea change is taking place in front of my eyes. After our software engineer son went abroad to execute a project for his company, she picked up chatting with him on Instant Messengers and was glad to see him 'live' on the Webcam. Now, every day, she is learning something new and singing hosannas to the Internet.

Perhaps the novelty will wear off after some time, but she has definitely gotten over her computer phobia. According to her, many of her friends are learning to use the Net.

Based on this observation, I have concluded that the Internet is going to see a burst of new users from India. I am certain that if all the initiatives that are being taken privately and publicly on bridging the Digital Divide between those who have access to the Net and those who do not are pursued seriously, then we might have over 200 million Internet users in India alone in ten to fifteen years.

That is a bold prediction considering that there are hardly 10 million PCs at the moment and 40 million telephone lines and the estimate of Internet users varies. Widely different figures from 3 million to 10 million are quoted. Nobody is sure. Like newspapers and telephones, Internet accounts too are heavily shared in India. In offices and homes, several people share single Internet account. And then there are cyber cafes too.

The Internet has become a massive labyrinthine library, where one can search for and obtain information. It has also evolved into an instant, inexpensive communication medium where one can send email and even images, sounds and videos, to a receiver, girdling the globe.

There are billions of documents in the Internet, on millions of computers known as Internet servers, all interconnected by a tangled web of cables, optic fibres and wireless links. We can be part of the Net through our own PC, laptop, cell phone or a palm-held Personal Digital Assistant, using a wired or a wireless connection to an Internet Service Provider. There are already hundreds of millions of users of the Internet.

You might have noticed that I have refrained from quoting a definite figure in the para above and have, instead, used ballpark figures. The reason is simple: I can't. The numbers are constantly changing even as we quote them. Like Jack's beanstalk, the Net is growing at a tremendous speed.

However, one thing we learn from 'Jack and the Beanstalk' is that every giant magical tree has humble origins. The beans, in the case of Internet, were sown as far back as the sixties.

It all started with the Advanced Research Projects Agency (ARPA) of the US Department of Defence. ARPA was funding advanced computer science research from the early '60s. J.C.R

Licklider, who was then working in ARPA, took the initiative in encouraging several academic groups in the US to work on interactive computing and time-sharing. We saw the historical importance of these initiatives in the chapter on computing.

One glitch, however, was that these different groups could not communicate their programmes or data or even ideas with each other easily. The situation was so bad that Taylor had three different terminals in his office in the Pentagon connected to three different computers that were being used for time -sharing experiments at MIT, UCLA and Stanford Research Institute. Thus started an experiment in enabling computers to exchange files among themselves. Bob Taylor played a crucial role in Information Processing Technology Office of ARPA in creating this network, which was later named Arpanet. "We wanted to create a network to support the formation of a community of shared interests among computer scientists and that was the origin of the Arpanet", says Taylor.

ARPANET WAS FOR COMMUNICATION

What about the story that the Arpanet was created by the US government's Defence Department, to have a command and control structure to survive a nuclear war? "That is only a story, and not a fact. Charlie Herzfeld, who was my boss at ARPA at one time, and I have made several attempts to clarify this. We should know, since we initiated it and funded Arpanet," says Taylor.

Incidentally, the two still remain good friends. When US president Bill Clinton awarded the National Technology Medal to Bob Taylor in 2000 for playing a leading role in personal computing and computer networking, Charles Herzfeld received the award on his behalf, since Taylor refused to travel to Washington, D.C.

It is a fact, however, that the first computer network to be proposed theoretically was for military purposes. It was to decentralize nuclear missile command and control. The idea was not to have centralized, computer-based command facilities, which could be destroyed in a missile attack. In order to survive a missile attack and retain what was known, during the US-Soviet Cold War, as 'Second Strike Capability', Paul Baran of Rand Corporation had proposed the idea of a distributed network. In those mad days of Mutually Assured Destruction, it seemed logical.

Baran elaborated his ideas to the military in an eleven-volume report 'Distributed Communications System' during 1962-64. This report was available to civilian research groups as well. However, no civilian network was built based on it. Baran even worked out the details of a packet switched network, though he used a clumsy name, 'Distributed Adaptive Message Block Switching'. Donald Davies, in the UK, independently discovered the same a little later and called it packet switching.

"We looked at Baran's work after we started working on the Arpanet proposal in February 1966", says Taylor. "By that time, we had also discovered that Don Davies in the UK had also independently proposed the use of packet switching for building computer networks. As far as we were concerned, the two papers acted as a confirmation that packet switching was the right technology to develop the ARPA network. That is all. The purpose of our network was communication and not ballistic missile defence", asserted Taylor passionately to the author. After his eventful innings at ARPA and Xerox PARC, Taylor is now retired to a wooded area not too far from the frenetic Silicon Valley.

DECENTRALISE

There was still a problem. How do you make Earthlings talk to Martians? Just kidding! Don't worry, I am trying to explain the difficulty in making two computers communicate with each other when two different manufacturers with different operating systems and software build them, as was done then. Without any exaggeration, the differences between ARPA computers at MIT and UCLA or Stanford Research Institute were as vast as that between Earthlings and Martians and assorted aliens.

"The problem was solved brilliantly by Wes Clark", says Bob Taylor. "He said let us build special-purpose computers to handle the packets, one at each 'host' (as each ARPA computer was known at that time). These special computers known as Interface Message Processors (IMPs) would be connected through leased telephone lines. ARPA would ask a contractor to develop the communication software and hardware for the IMPs while each site would worry about developing the software to make their host computer talk to the local IMP. Since the research group at each site knew their own computer inside out, they would be able to do it. Moreover, this approach involved many scattered research groups in building the network and not remain as were users of a network", says Taylor, well-known as a motivator and a subtle manager. Thus, the actual network communication through packets took place between standardized IMPs designed centrally.

At one stroke, Wesley Clark had solved the problem of connecting Earthlings to Martians using IMPs. As it turned out, Donald Davies had also arrived at the same conclusion in the UK but could not carry it further since the UK did not have a computer networking project at that time.

As a management case study, the execution of Arpanet is very interesting. A path-breaking initiative involving diverse elements is complex to execute. The skill of management lies in separating global complexity from local and centralising global complexity while decentralizing the resolution of local complexity. The execution of Arpanet was one such. Mature management was as important as the development of new technology for its speedy and successful build-up.

POST OFFICES AND PACKET SWITCHING

What is packet switching? It is the single most important idea in all computer networks, be it the office network or the global Internet. The idea is simple. In telephone networks (as we saw in the chapter on telecommunication), the two communicating parties are provided a physical connection for the duration of the call. The switches and exchanges take care of that. Hence, this mode of establishing a communication link is called 'circuit switching'. If, for some reason, the circuit is cut due to the failure of a switch or the transmission line getting cut or something similar, then the call gets cut too.

In advanced telephone networks, your call may be routed through some other trunk line, avoiding the damaged section. However, this can take place only in a second try and if an alternative path is available. Moreover, a physical connection is again brought into being for the duration of the call.

In packet switching, however, if the computer is trying to send a 'file' (data, programs or email) to another computer, then it is first broken into small packets. The packets are then put inside 'envelopes' with the address of the computer where they are supposed to go along with the 'sequential number' of the packet. These packets are then let loose in a network of packet switches called Routers.

Each router, which receives the packet, acts like a sorter in a post office who reads the address, does not open the envelope and sends it on its way to the right destination. Thus the packet reaches the destination, where it is used in the appropriate order to reassemble the original message.

Of course, there is an important difference between the postal sorter and an Internet packet switch or router. A postal sorter in Mumbai will send a letter addressed to Kumbhakonam in a big bag to Chennai along with all the letters going to Tamil Nadu. The Chennai sorter will then send it to Kumbhakonam. The sorter in Mumbai will never (we hope!) send it to Kolkata. But a network router might do just that.

The router will see if the path to Chennai is free; if not, it will send it to the router in Kolkata, which will read the Kumbhakonam address; but if the Kolkata-Chennai path is not free, then it might send it to Bangalore; and so on. In short, routers have the intelligence to sense the traffic conditions in the network, continuously update themselves on the state of the network, decide on the best path at the moment for the packet and send it forward accordingly.

This is very different from the method used by postal sorter. What if congestion at Chennai leads to packet loss? Moreover, what if a link in the chain is broken? In the case of the postal system, it might take a long time to restore the service, but in the case of packet switching, if a router goes down, then it does not matter—the network will find some other path to send the packet. Despite all this, if the packet does not reach the destination, then the computer at the destination will ask for it to be sent again.

You might say this does not look like a very intelligent way of communicating, and you would not be alone; the whole telecom industry said so. Computer networking through packet switching was opposed tooth and nail by telecom companies and even those with cutting-edge technologies like AT&T! They suggested that leased lines be used to connect computers and that was that.

PACK IT UP SAID AT&T

Networking pioneers like Paul Baran, Bob Taylor, Larry Roberts, Frank Heart,Vinton Cerf, Steve Crocker, Bob Metcalfe, Len Kleinrock, Bob Kahn and others have recalled, in several interviews, the struggle they had to go through to convince AT&T, the US telephone monopoly of those days.

AT&T did not believe packet switching would work, and that, if it ever did, it would become a competing network and kill their business! This battle between data communication and incumbent telephone companies is still not over. As voice communication adopts packet technology, as in Voice Over Internet, the old phone companies all over the world are barely conceding to packet switching, kicking and crying.

This may look antediluvian, but forty years ago, the disbelievers did have a justification: the computers required to route packets were neither cheap enough nor fast enough.

The semiconductor and computer revolution has taken care of that and today's routers look like sleek DVD or VCR players and cost a fraction of old computers. Routers are actually very fast, special-purpose computers; hence packets take microseconds to be routed at each node.

The final receiver is able to reassemble the message with all the packets intact and in the right order in a matter of a few milliseconds. Nobody is the wiser about what goes on behind the scenes with packets zigzagging through the network madly before reaching the destination.

In the case of data communication, making sure that none of the packets have been lost, and not necessarily the time taken, becomes more important. For example, if I send the publisher this chapter through the network, I do not want several words, sentences or punctuations missing or jumbled up taking the Mickey out of it. However, if I am sending a voice signal in packetised form, then the receiver is interested in real-time communication even at the cost of a few packets. That is the reason the voice appears broken when we use Instant Messengers to talk a friend on the Net. However, with constant R&D, the technology of packet switching is improving and one can carry out a decent voice or even video communication or listen to radio or see a TV broadcast on the Internet.

WHY NOT CIRCUIT SWITCHING?

The main advantage of using packet switching in computer communication is that it uses the network most efficiently. That is why it is also the most cost-effective. We have already seen the effort telecom engineers make to use their resources optimally through clever multiplexing and circuit switching; so why don't we just use telephone switches between computers?

There is a major difference between voice and data communication. Voice communication is more or less continuous, with a few pauses, but computer communication is bursty. That is, a computer will send megabytes for some time and then fall silent for a long time, and we would be blocking a telephone line and that, too, mostly long-distance lines, thereby making it very expensive.

Suppose you have a website on a server and I am visiting the website. I find a file interesting and I ask for it through my browser. The request is transmitted to your server and it sends the file in a few seconds. Then I take fifteen minutes to read that file before making another request. Imagine what would happen if I were in Mumbai and your server was in Milwaukee and we were circuit switched? An international line between Mumbai and Milwaukee has to be kept open for fifteen minutes, waiting for the next request! If the Internet were based on circuit switching, then not only would it be expensive but also just a few lakh users would tie up the entire global telephone networks.

Using ARPA funds, the first computer network based on packet switching was built in the US between 1966 and 1972. A whole community of users came into being at over a dozen sites, and started exchanging files. Soon they also developed a system to exchange notes and they called it 'e-mail' (an abbreviation for electronic mail). Abhay Bhushan, who worked in the Arpanet project from 1967 to 1974 was then at MIT and wrote the note on FTP or File Transfer Protocol, the basis of email. In those days, several theoretical and practical problems were sorted out through RFCs, which stood for Request For Comments –a message sent to all Arpanet users. Any researcher in a dozen ARPA sites could pose a problem or post a solution through such RFCs. Thus, an informal, non-hierarchical culture developed among these original Netizens. "Those were heady days when so many things were done for the first time without much ado," recalls Abhay Bhushan.

WHO PUT @ IN EMAIL?

A form of email was already known to users of time-sharing computers, but with Arpanet coming into being, new programs started popping up for sending mail piggyback on File Transfer Protocol. An email program that immediately became popular due to its simplicity was sendmsg, written by Ray Tomlinson, a young engineer at Bolt Beranek and Newman (BBN), a Boston-based company, which was the prime contractor for building the Arpanet. He also wrote a program called readmail to open the email. His email programs have obviously been superseded in the last thirty years by others. But one thing that has survived is the @ sign to denote the computer address of a sender. Tomlinson was looking for a symbol to separate the receiver's user name and the address of his host computer. When he looked at his Teletype, he saw a few punctuation marks available and chose @ since it had the connotation of 'at' among accountants, and did not occur in software programs in some other connotation.

An idea some Arpanet pioneers like Larry Roberts were promoting in those days to justify funding the project was that it would be a resource sharing project. That is, if your ARPA computer is overloaded with work and there is another ARPA computer across the country, which has some free time, then you could connect to that computer through the Arpanet and use it. Though it sounded reasonable, it never really worked that way. Computer networks, in fact, came to be increasingly used for communication of both the professional and personal kind. Today, computer-to-computer communication through email and chat has become one of the 'killer apps'—an industry term for an application that makes a technology hugely popular and hence provides for its sustenance—for the Internet.

LOCAL AREA NETWORKS

The Arpanet matured during the '70s. Bob Taylor who had left Arpa in 1969 had started the computing research division at brand new Xerox Parc, at Palo Alto, California. His inspiration remained Licklider's dream of interactive computing. At PARC it evolved into an epoch-making project of personal computing that led to the development of mouse, icons and graphical user interfaces, windows, laser printer, desktop computer and so on, which we have discussed in the chapter on personal computing. Parc scientists were also the first users of their own technology. Altos was developed at PARC as a desktop machine for its engineers.

For Taylor, connecting these desktop computers in his lab was but a natural thing. So he assigned Bob Metcalfe, a brilliant engineer, who had earlier worked with Abhay Bhushan at MIT on the Arpanet, to the task. Interestingly, both Metcalfe and Bhushan shared a dislike for academic snobbery. MIT, the Mecca of engineering, had turned down Bhushan's work on Arpanet as being too low-brow for a PhD, while Harvard had turned down Metcalfe's. "They probably wanted lots of equations and Greek symbols," said Metcalfe once, sarcastically.

As a footnote, it is worth noting that, later, Harvard accepted Metcalfe's analysis of the Alohanet experiment in Hawaii for a PhD 'reluctantly', according to Metcalfe. MIT, however, has made up for Harvard by putting Metcalfe on its governing board.

What was Alohanet? When Metcalfe started experimenting on Local Area Network (LAN) at PARC, he looked at Alohanet, which had already come into being in Hawaii, thanks to ARPA funding. Since Hawaii is a group of islands, the only way the computer at the University of Hawaii's main campus could be connected to the terminals at other campuses on different islands was through a wireless network. It was appropriately called Alohanet, since 'Aloha' is the Hawaiian equivalent of 'Hi'. Norman Abramson, a professor of communication engineering at the University of Hawaii, had designed it.

The Alohanet terminals talked through radio waves with an IMP, which then communicated with the main computer. It looked like a straightforward extension of Arpanet ideas. But there was a difference. In the Arpanet, leased telephone lines connected the IMPs, and each IMP could 'see' the traffic conditions and send its packets. In Alohanet, however, the terminals could not know in advance the packets that were floating in the Ether. So they would wait for the destination computer to acknowledge the receipt of the packet and if they did not get an acknowledgement then they would send the packet again, after waiting for a random amount of time. With a limited number of computers, the system worked well.

Metcalfe saw similarities between Alohanet and his local area network (LAN) inside Xerox PARC, in that each computer sent its packets into the void or ether of a coaxial cable and hoped that the packets reached the destination. If they did not get an acknowledgement then they would wait randomly a few microseconds and send the packets again. The only difference was that, while Alohanet was quite slow due to low bandwidth, Metcalfe could easily jack up the speed of his LAN to a few megabits per second. He named his network protocol and the architecture as Ethernet.

WHAT IS A NETWORK PROTOCOL?

A 'communication protocol' is a favourite word of networking engineers just as 'algorithm' is a favourite of computer scientists. Leaving the technical details aside, a protocol is actually a step-by-step approach to enable two computers "talk to each other" i.e. exchange data. We use protocols all the time in human communication, so we don't notice it, but if two strangers met, then how would they start to converse? They would start by introducing themselves, finding a common language, agreeing on a level of communication—formal, informal, professional, personal, polite, polemical and so on, before exchanging information.

Computers do the same thing through different protocols. For example, what characterises Alohanet and Ethernet protocols is that packets are sent again after a random wait if they were lost due to data collision. We also do it so often. If two people start talking at once, then their information would 'collide' and not reach the other person—the 'destination'. They would then wait politely for a random amount of time, for the other person to start talking and only if he or she does not, then they would. That is what computers do too, if connected by Ethernet.

When Xerox, Intel and DEC agreed to adopt Ethernet as a networking standard and made it public in 1980, Metcalfe saw an opportunity and started a company called 3COM (Computers, Communications, Compatibility) to supply Ethernet cards and other equipment. Within a few years 3COM had captured the office networking market even though other proprietary systems from Xerox and IBM were around. This was mainly because Ethernet had become an open standard and 3COM could network any set of office computers and not just those made by IBM or Xerox. Thereby hangs another tale of the fall of proprietary technology and the super success of open standards that give the option to the user to choose his components.

Today, using fibre optics, Ethernet can deliver lightning data speeds of a 100 Mbps. In fact, a new version of Ethernet called Gigabit Ethernet is fast becoming popular as a technology to deliver high-speed Internet to homes and offices.

Local Area Networks have changed work habits in offices across the globe like never before. They have led to file-sharing, smooth workflow and collaboration, besides speedy communication within companies and academic campuses.

THE PENTAGON AS THE CHESHIRE CAT

As Arpanet rose in popularity in the 70s, a clamour started from every university and research institution to be connected to Arpanet. Everybody wanted to be part of this new community of shared interests. However, not everyone in a Local Area Network could be given a separate Arpanet connection, so one needed to connect entire LANs to Arpanet. Here again there was a diversity of networks and protocols. So how would you build a network of networks (also called the Internet)? Largely, Robert Kahn and Vinton Cerf solved this problem by developing TCP (Transmission Control Protocol) and hence they are justly called the inventors of the Internet.

Regarding the motivation for the Internet, Cerf pointed out that one of them was definitely defence. "Arpanet was built for civilian computer research, but when we looked at connecting various networks we found that there were experiments going on packet satellite networks to network naval ships and the shore installations using satellites. Then there were packet radio networks, where packets were sent wireless by mobile computers. Obviously, the army was interested, since it represented to them some battlefield conditions on the ground. In fact, today's GPRS (General Packet Radio Service) or 2.5 G cell phone networks are based on the old packet radio. At one time, we put packet radios on air force planes, too, to see if strategic bombers could communicate with each other in the event of nuclear war. So the general problem was how to internetwork all these different networks, but the development of technology had obvious military interest," says Cerf. "But even the highway system in the US was built with missile defence as a justification. It lead to a leap in the automobile industry, in housing, and changed the way we live and work besides transportation of goods, but I do not think any missile was ever moved over a highway," adds he.

Coming back to the problem of building a network of networks, Cerf says, "We had the Network Control Protocol to run the Arpanet, Steve Crocker had led that work. But the problem was that NCP assumed that packets would not be lost, which was okay to an extent within Arpanet, but Bob Kahn and I could not assume the same in the Internet. Here, each network was independent and there was no guarantee that packets would not be lost, so we needed recovery at the edge of the net. When we first wrote the paper in 1973-74 we had a single protocol called TCP. Routers could take packets that were encapsulated in the outer networks and carry it through the Internet. It took four iterations from 1974-78 to arrive at what we have today. We split the TCP program into two. One part worried about just carrying packets through the multiple networks while the other part worried about restoring the sequencing and looking at packet losses. The first was called IP (Internet Protocol) and the other, which looked at reliability, was called TCP. Therefore, we called the protocol suite, TCP/IP. Interestingly, one of the motivations for separating the two was to carry speech over the Internet," reveals Cerf.

TCP allowed different networks to get connected to the Arpanet. The IMPs were now divided into three boxes: one dealt with packets going in and out of the LAN, the other dealt with packets going in and out of the Arpanet and a third, called a gateway, passed packets from one IMP to the other while correctly translating them into the right protocols. Meanwhile, in 1971, an undergraduate student at IIT Bombay, Yogen Dalal, was frustrated by the interminable wait to get his programs executed by the old Russian computer. Thanks to encouragement from a faculty member, J R Isaac, who was then head of the computer centre, Dalal started a BTech project on building a remote terminal for the mainframe. "Like all undergraduate projects, this also did not work," laughs Dalal, recalling those days. But when he went to Stanford for his MS and PhD and saw cutting-edge work being done in networking by Cerf & Co., he naturally got drawn into it.

As a result, Vinton Cerf, Yogen Dalal and another graduate student, Carl Sunshine, wrote the first paper setting forth the standards for an improved version of TCP/IP, in 1974, which became the standard for the Internet. "Yogen did some fundamental work on TCP/IP. I remember, during 1974, when we were trying to sort out various problems of the protocol, we would come to some conclusions at the end of the day and Yogen would go home and come back in the morning with counter examples. He was always blowing up our ideas to make this work," recalls Cerf.

"They were the most exciting years of my life," says Yogen Dalal, who after a successful career at Xerox PARC and Apple, is a respected venture capitalist in Silicon Valley. Recently he was listed as among the top fifty venture capitalists in the world.

THE TANGLED WEB

In the eighties, networking expanded further among academics, and the Internet evolved as a communication medium with all the trappings of a counter culture.

Two things changed the Internet, meant for the specialist to the Internet that millions could relate to. One was the development of the World Wide Web and the other was a small program called the Browser that allowed you to navigate in this web and read the web pages.

The web is made up of host computers connected to the Internet containing a program called a Web Server. The Web Server is a piece of computer software that can respond to a browser's request for a page and deliver the page to the Web browser through the Internet. You can think of a Web server as an apartment complex with each apartment housing someone's Web page. In order to store your page in the complex, you need to pay rent on the space. Pages that live in this complex can be displayed to and viewed by anyone all over the world. The host computer is your landlord and your rent is called your hosting charge. Every day, there are millions of Web servers delivering pages to the browsers of tens of millions of people through the network we call the Internet.

The host computers connected to the Net, called Internet servers, are given a certain address. The partitions within the server hosting separate documents belonging to different owners are called Websites. Each website in turn is also given an address—Universal Resource Locator (URL). These addresses are assigned by an independent agency. It acts in a manner similar to that of the registrar of newspapers and periodicals or the registrar of trade marks, who allow you to use a unique name for your publication or product if others are not using it.

When you type in the address or URL of a website in the space for the address in your browser, the program sends packets requesting to see the website. The welcome page of the website is called the home page. The home page carries an index of other pages, which are part of the same website and residing in the same server. When you click with your mouse on one of them, the browser recognises your desire to see the new document and sends a request to the new address, based on the hyperlink. Thus, the browser helps you navigate the Web or surf the information waves of the Web—which is also called Cyberspace, to differentiate from real navigation in real space.

The web pages carry composing or formatting instructions in a computer language known as Hyper Text Markup Language (HTML). The browser reads these instructions or tags when it displays the web page on your screen. It is important to note that the page, on the Internet, does not actually look the way it does on your screen. It is a text file with embedded HTML tags giving instructions like 'this line should be bold', 'that line should be in italics', 'this heading should be in this colour and font,' 'here you should place a particular picture' and so on. When you ask for that page, the browser brings it from the Internet web servers and displays it according to the coded instructions. A web browser is a computer program in your computer that has a communication function and a display function. When you ask it to go to an Internet address and get a particular page, it will send a message through the Internet to that server and get the file and then, interpreting the coded HTML instructions in that page, compose the page and display it to you.

An important feature of the web pages is that they carry hyperlinks. Such text (with embedded hyperlinks) is called Hyper Text, which is basically text within text. For example, in the above paragraphs, there are words like 'HTML', 'World Wide Web' and 'Browser'. Now if these words are hyperlinked and you want to know more about them, then I need not give the information right here, but provide a link to a separate document to explain each of these words. So, only if you want to know more about them, would you go that deep.

In case you do want to know more about the Web and you click on it, then a new document that appears might explain what the Web is and how it was invented by Tim Berners-Lee, a particle physicist, when he was at CERN, the European Centre for Nuclear Research at Geneva. Now if you wanted to know more about Tim Berners-Lee or CERN then you could click on those words with your mouse and a small program would hyperlink the words to other documents containing details about Lee or CERN and so on.

Thus, starting with one page, you might 'crawl' to different documents in different servers over the Net depending on where the hyperlinks are pointing. This crawling and connectedness of documents through hyperlinks seems like a spider crawling over its web and there lies the origin of the term 'World Wide Web.'

STORY WITHIN A STORY

For a literary person, the hyperlinked text looks similar to what writers call non-linear text. A linear text has a plot and a beginning, a middle and an end. It has a certain chronology and structure. But a nonlinear text need not have a beginning, middle and an end in the normal sense. It need not be chronological. It can have flashbacks and flash-forwards and so on.

If you were familiar with Indian epics then you would understand hyperlinked text right away. After all, *Mahabharat,*[1] *Ramayana,*[2] *Kathasaritsagar,*[3] *Panchatantra,*[4] Vikram and Betal's[5] stories have nonlinearities built into them. Every story has a sub-story. Sometimes there are storytellers as characters within stories, who then tell other stories, and so on. At times you can lose the thread because, unlike Hyper Text and hyperlinks—where the reader can exercise his choice to follow a hyperlink or not—the sub-stories in our epics drag you there anyway!

Earlier, you could get only text documents on the Net. With HTML pages, one could now get text with pictures or animations or even some music clips or video clips and so on. The documents on the Net became so much livelier, while the hyperlinks embedded within the page took you to different servers—host computers on the Internet acting as repositories of documents.

It is as if you open one book in a library and it offers you the chance to browse through the whole library of books, CDs and videos! By the way, the reference to the Web as a magical library is not fortuitous. This idea of a hyperlinked electronic library was essentially visualised in the 1940s by Vannevar Bush at MIT, which he had called Memex.

Incidentally, Tim Berners-Lee was actually trying to solve the problem of documentation and knowledge management in CERN. He was grappling with the problem of how to create a database of knowledge so that the experience of the past could be distilled in a complex organisation. It would also allow different groups in a large organisation to share their knowledge resources. That is why his proposal to his boss to create a hyperlinked web of knowledge within CERN, written in 1989-90, was called: 'Information Management: A Proposal'. Luckily, his boss is supposed to have written two famous words, "Why not?" on his proposal. Lee saw that the concept could be generalised to the Internet. The Internet community quickly grasped it, and we saw the birth of the Internet as we know it today. A new era had begun.

Lee himself developed a program, that looked like a word processor and had hyperlinks as underlined words. He called it a browser. The browser had two functions: a communication function which used Hyper Text Transfer Protocol (HTTP) to communicate with servers, and a presentation function. As more and more servers capable of using HTTP were set up, the Web grew.

Soon more browsers started appearing. The one written by a graduate student at the University of Illinois, Marc Andreessen, became very popular for its high quality and free downloading. It was called Mosaic. Soon, Andreessen left the university, teamed up with Jim Clark, founder of Silicon Graphics, and floated a new company called Netscape Communications. Its Netscape Navigator created a storm and the company started the Internet mania on the stock market when it went public, attracting billions of dollars in valuation even though it was not making any profit!

---

[1] *India's greatest epic, based on ancient Sanskrit verse, of sibling rivalry.*

[2] *Ancient Indian epic—The Story of Rama.*

[3] *Ancient collection of stories.*

[4] *Anonymous collection of ancient animal fables in Sanskrit.*

[5] *A collection of 25 stories where a series of riddles are posed to king Vikram by Betal, a spirit.*

Meanwhile, Tim Berners-Lee did not make a cent from his path breaking work since he refused to patent it. He continues to look at the development of the next generation of the Internet as a non-profit service to society and heads a research group, W3C, at MIT, which has become a standards-setting consortium for the Web.

With the enormous increase in the number of servers connected to the net carrying millions of documents, the need also arose of efficiently searching them. There were programs to search databases and documents. But how do you search the whole Web? Thus, programs were written to collect a database of key words in Internet documents and they are known as search engines. They list the results of such searches in the order of frequency of occurrence of the keywords in different documents. Thus, if I am looking for a document written by 'Tim Berners-Lee' on the Web, then I type the words 'Tim Berners-Lee' in the search engine and ask it to search the web for it. Within a few seconds, I get a list of documents on the web with the words Tim Berners-Lee.' They could have been written by Lee or about Lee. HTML documents carry keywords used in the document, called meta-tags. Initially, it was enough to search within the Meta-tags, but now powerful search engines have been devised which search the entire document. They make a list of words which are really important, based on the frequency of occurrence and the place where they appear. That is, do they occur in the title of the document or in subheadings or elsewhere? They then assign different levels of importance to all these factors, which is a process called 'weighting'. Based on the sum of weights, they rank different pages and then display the search results.

LIBRARIAN FOR THE INTERNET

Before the Web became the most visible part of the Internet, there were search engines in place to help people find information on the Net. Programs with names like 'gopher' (sounds like 'Go For') and 'Archie' kept indexes of files stored on servers connected to the Internet and dramatically reduced the amount of time required to find programs and documents.

A web search engine employs special autonomic software called spiders to build lists of words found on Web sites. When a spider is building its lists, the process is called Web crawling. In order to build and maintain a useful list of words, a search engine's spiders have to look at many pages.

How does a spider crawl over the Web? The usual starting points are lists of heavily used servers and very popular pages. The spider will begin with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel, spreading out across the most widely used portions of the Web.

In the late nineties, a new type of search engine was launched by two graduate students in Stanford, viz. Larry Page and Sergey Brin called Google. It goes beyond keyword searches and looks for 'connectedness' of the document and has become the most popular search engine at the moment.

Rajeev Motwani, a professor of computer science at Stanford, encouraged these students by skunking away research funds to them to buy new servers for their project. He explains, "Let us say that you wanted information on 'bread yeast' and put those two words in Google. Then it not only sees which documents have these as words mentioned but also whether these documents are linked to other documents. An important page for 'bread yeast' must have all other pages on the web dealing in any way with 'bread yeast' also linking to it. In our example, there may be a Bakers' Association of America, which is hyperlinked by most documents containing 'bread yeast', implying that most people involved with 'bread' and 'yeast' think that the Bakers Association's web site is an important source of information. So Google will then rate that website very high and put it on top of its list. Thus irrelevant documents which just mention 'bread' and 'yeast' will not be given any priority in the results."

Motwani who is a winner of the prestigious Gödel Prize for his contributions to computer science, is a technical advisor to Google and is watching its growth with enthusiasm. Google today boasts of being the fastest search engine and even lists the time taken by it (usually a fraction of a second) to scan billions of documents to provide you with results. The strange name Google came about because 'googol' is a term for a very large number (10 to the power 100 coined by a teenager in America in 1940s) and the founders of Google, who wanted to express the power of their new search engine, misspelt it!

By the way, you might have noticed that the job of the search engine is nothing more than what a humble librarian does all the time and more intelligently! However, the automation in the software comes to our rescue in coping with the exponential rise in information.

NEW MEDIA

With the user-friendliness of the Internet taking a quantum leap with the appearance of the Web, commercial interests took notice of its potential. What was until then a communication medium for engineers now appeared accessible to ordinary souls that were looking for information and communication. That is what every publication does and hence the Web was looked upon as a new publishing medium.

Now, if there is a new medium of information and communication and if millions of people are 'reading' it, then will advertising be far? News services and all kinds of publications started using the Web to disseminate news like a free wire service.

An ordinary person too could put up a personal web page at a modest cost or for free using some hosting services containing his personal information or writings. Thus, if Desktop Publishing created thousands of publications, the Web led to millions of publishers!

As a medium of communication, corporations and organisations all over the world have adopted Web technology. A private network based on web technology is called an Intranet, as opposed to the public Internet. Thus, besides the local area networks, a corporate CEO has a new way to communicate with his staff. But progressive corporations are using the reverse flow of communication through their Intranets—from the bottom up as well, to break rigid bureaucracies and 'proper channels' to rejuvenate themselves.

The Web, however, was more powerful than all old media. It was interactive. The reader can not only read what is presented but can send requests for more information about the goods and services advertised.

Meanwhile, developments in software led to the possibility of forms being filled and sent by users on the Web. Web pages became dynamic, one could see 'buttons' and even a click was heard when you clicked on the button! Dynamic HTML and then Java enriched the content of Web pages. Java was a new language, developed by Sun Microsystems, that could work on any operating system. It was ideally suited for the Web, since no one knew the variety of hardware and operating systems inside the millions of computers that were connected to the Web. Interestingly, when I asked Raj Parekh, who worked as VP Engineering and CTO of Sun, about how the name Java was picked, he said, "We chose Java because Java beans yield strong coffee, which is popular among engineers". Today the steaming coffee cup—the symbol of Java—is well known to software engineers.

Developments in software now led to encryption of information sent by the user to the web server. This provided the possibility of actually transacting business on the web. After all, any commercial transaction needs security.

Computers containing the personal financial information of the user, like those of banks and credit card companies, could now be connected to the Web with appropriate security like passwords and encryption. At one stroke, the user's request for information regarding a product could be turned into a 'buy order' his bank or credit card company being duly informed of the same. Thus a wave of a new type of commerce based on the Web came into being, called e-commerce. The computers that interfaced the Web with the bank's database came to be known as payment gateways.

New enterprises that facilitated commerce on the Net were called dotcoms. Some of them actually sold goods and services on the Net, while others only supplied information. The information could also be a commodity to be bought or distributed freely. For example, if you wanted to know the details of patents filed in a particular area of technology then the person who had digitised the information, classified it properly and made it available on the web might charge you for providing that information, whereas a news provider may not charge any money from visitors to his website and might collect the same from advertisers instead.

COME, SET UP A SHOP WINDOW

There were companies like Amazon.com, which started selling books on the Internet. This form of commerce still needed warehouses and shipping and delivery of goods once ordered but they saved the cost of the real estate involved in setting up a retail store. It also made the 'store front' available to anybody on the Net, no matter where he was sitting.

E-commerce soon spread to a whole lot of retail selling, be it airline or railway tickets or hotel rooms or even financial services. Thus, one could sit at home and book a ticket or a hotel across the world or have access to his bank account. You could not only check your bank account but also now pay bills on the Web—your credit card bills, telephone, electricity and mobile bills, etc thereby minimising your physical visits to various counters in diverse offices.

With the web bringing the buyer and seller together directly, many transactions that had to go through intermediaries could now be done directly. For example, one could auction anything on the web as long as both the parties trusted each other's ability to deliver. If you could auction your old PC or buy a used car, then you could put up your shares in a company to auction as well.

But that is what stock exchanges do. So computerised stock exchanges, like Nasdaq in the US and NSE in India, which had already brought in electronic trading, could now be linked to the web. By opening your accounts with certain brokers, you could directly trade on the Net, without asking your broker to do it for you.

In fact, Web market places called exchanges came into being to buy and sell goods and services for businesses and consumers. At one time, a lot of hype was created in the stock markets of the world that this form of trading would supersede all the old forms and a 'new economy' has come into being. Clearly, it was an idea whose time had not yet come. The Internet infrastructure was weak. There was a proliferation of web storefronts with 'no ware houses and goods or systems of delivery' in place. The consumers balked at this new form of catalogue marketing and even businesses clearly showed preferences for trusted vendors. But the web technologies have survived. Internet banking is growing and so is bill payment. Even corporations are linking their computer systems with those of their vendors and dealers in a Web-like network for collaboration and commerce. Services like booking railway or airline tickets or hotel rooms are being increasingly used. The web has also made it possible for owners of these services to auction a part of their capacity to optimise their occupancy rate.

Some entrepreneurs have gone ahead and created exchanges where one can look for a date as well!

Clearly, we are going to see more and more transactions shifting to the Internet, as governments, businesses and consumers all get caught up in this magical Web.

THEN THERE WAS HOTMAIL

It might sound like an Old Testament-style announcement, but that cannot be helped because the arrival and growth of email has changed communication forever.

In the early nineties, Internet email programs existed on a limited scale and one had to pay for them. Members of academia or people working in large corporations had email facility but those outside these circles could not adopt it unless they subscribed to commercial services like America On Line. Two young engineers, Sabeer Bhatia and Jack Smith thought there must be a way for providing a free email service for anybody who registers at their website. One could then access the mail by just visiting their website from anywhere. This idea of web-based mail, which was named Hotmail, immediately caught the fancy of millions of people. Microsoft acquired the company. Meanwhile, free Web-based email services have played a great role in popularising the new communication culture and, today, Hotmail would be one of the largest brands on the Internet.

Soon, 'portals' like Yahoo offered web mail services. A portal is a giant aggregator of information, and a catalogue of documents catering to varied interests. Thus, a Web surfer could go to one major portal and get most of the information he wanted through the hyperlinks provided there. Soon, portals provided all kinds of directory services like telephone numbers.

As portals tried to be one-stop shops in terms of information, more and more directory services were added. Infospace, a company founded by Naveen Jain in Seattle, pioneered providing such directory services to websites and mobile phones and overwhelmingly dominates that market in the US.

BLIND MEN AND THE ELEPHANT

The Web has become many things to many people.

For people like me looking for information, it has become a library. "How about taking this further and building giant public libraries on the Internet," asks Raj Reddy. Reddy, a veteran computer scientist involved in many pioneering projects in Artificial Intelligence in the '60s and '70s, is now involved in a fantastic initiative of scanning thousands of books and storing them digitally on the Net. It is called the Million Book Digital Library Project at Carnegie Mellon University. Reddy has been meeting with academics and government representatives in India and China to make this a reality. He is convinced that this labour-intensive job can best be done in India and China. Moreover, the two countries will also benefit by digitising some of their collections. A complex set of issues regarding Intellectual Property Rights, like copyright, are being sorted out. Reddy is hopeful that pragmatic solutions will be found for copyright issues.

Public libraries in themselves have existed for a long time and shown that many people sharing a book or a journal is in the interest of society.

When universities across the world and especially in India are starved of funds for buying new books and journals and housing them in decent libraries, what would happen if such a Universal Digital Library is built and made universally available? That is what N Balakrishnan, professor at Indian Institute of Science, Bangalore, said to himself. He then went to work with a dedicated team to lead the work of the Million Book Digital Library in India. Already, several institutions are collaborating in this project and over fifteen million pages of information have been digitized from books and journals and even palm leaf manuscripts.

If the governments intervene appropriately and the publishing industry cooperates, then there is no doubt that a Universal Digital Library can do wonders to democratise access to nowledge.

WHAT NEXT?

While Rajeev Motwani and his former students dream of taking over the world with Google, Tim Berners-Lee is evangelising the next generation of the Web, which is called the Semantic Web. In an article in *Scientific American,* May 2001, Lee explained, "Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. Computers can adeptly parse Web pages for layout and routine processing—here a header, there a link to another page—but in general, they have no reliable way to process the semantics. The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. It is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The initial steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines are enabled to process and 'understand' the data that they merely display at present.

"Information varies along many axes. One of these is the difference between information produced primarily for human consumption and that produced mainly for machines. At one end of the scale, we have everything from the five-second TV commercial to poetry. At the other end, we have databases, programs and sensor output. To date, the Web has developed most rapidly as a medium of documents for people rather than for data and information that can be processed automatically. The Semantic Web aims to make up for this. The Semantic Web will enable machines to comprehend semantic documents and data, not human speech and writings," he adds.

It is easy to be skeptical about a semantic web, as it smells of Artificial Intelligence, a project that proved too ambitious. Yet Lee is very hopeful of results flowing in slowly. He recognises that the Semantic Web will need a massive effort and is trying to win over more and more people to work on the challenge.

To sum up the attributes of the Net developed so far, we can say it has become a market-place, a library and a communication medium.

A MANY-SPLENDOURED THING

The dual communication properties of the Web—inexpensive broadcasting and interactivity—will lead to new applications:

- It will help intellectual workers work from homes, etc and not necessarily from offices, thereby reducing the daily commute in large cities. This is being called telecommuting.

- It can act as a great medium to bring the universities closer to people and encourage distance learning. Today, a distance learning experiment is in progress in the Adivasi villages of Jhabua district in Madhya Pradesh, using ISRO satellite infrastructure and a telephone link between students and the teacher for asking questions. Several such 'tele-classrooms' could come alive, for everyone, on their PC screens. Imagine students in remote corners of India being able to see streamed videos of the Feynman Lectures in Physics or a demonstration of a rare and expensive experiment or a surgical operation! With improved infrastructure, the ‘return path’ between students and teachers can be improved. Then a student need not go to an IIM or IIT to take a course, he could do it remotely and even send questions to the professor. It will increase the reach of higher education and training many-fold..

- It can act as a delivery medium for video on demand and music on demand, thus bringing in entertainment functions. Basavraj Pawate, a chip design expert at Texas Instruments who is now working on VOIP, believes that CD-quality sound can be delivered through the Net, once some problems of Internet infrastructure are solved.

- When video conferencing on the Net becomes affordable, all kinds of consultations with doctors, lawyers, accountants and so on can become possible at a distance. For example, today, a heart specialist in Bangalore is linked to villages in Orissa and Northeastern India. He is able to talk to the patients and local doctors, receive echocardiograms of patients and give his expert opinion. Similarly, the Online Telemedicine Research Institute of Ahmedabad could provide important support during Kumbh Mela[†] and in the aftermath of earthquakes like that in Bhuj.
However, all these experiments have been made possible using VSATS, thanks to ISRO offering the satellite channel for communication. As Internet infrastructure spreads far and wide, the same could be done between any doctor and any patient. Today, premier medical research institutions like All India Institute of Medical Sciences, Delhi, the Post Graduate Institute of Medical Sciences and Research, Chandigarh, and the Sanjay Gandhi Institute of Medical Sciences, Lucknow are being connected with a broadband network for collaborative medical research and consultation. However, if a broadband network comes into being in all the cities and semi- urban centres, to begin with, then medical resources concentrated in the big cities of India can be made available to patients in semi-urban and even rural areas. Thus an expert’s time and knowledge can be optimally shared with many people.

- Indian farmers have demonstrated their hunger for knowledge and new agri-technology in the last thirty years. The Net can be used to distribute agronomical information, consultation regarding pests and plant diseases, meteorological information, access to land records and even the latest prices in the agricultural commodity markets. Though India has a large population involved in agriculture, the productivity of Indian agriculture is one of the lowest in the world. The Net can thus be used to increase agricultural productivity by enabling a massive expansion of the extension programme of agricultural universities and research institutions. Already, several experiments are going on in this direction. However, large-scale deployment of rural Internet kiosks, akin to the ubiquitous STD booths, awaits large-scale rural connectivity.

- It is understood that Voice Over Internet, that is, packetised voice, might become an important form of normal telephony. International telephony between India and US has come down drastically in cost due to VOIP. When this technology is applied within India, a cheaper and more affordable telephony can be provided, which will work wonders with the Indian economy. A large number of the poor in the cities, small towns and villages will be brought into the telecom net. While the incumbent state- owned telephone company, BSNL, might take some time to adopt new IPbased

---

[†] *A large gathering of Hindu devotees to take a sacred dip in the holy rivers.*

(Internet Protocol) technologies for voice due to legacy issues in its network, the new private networks have an opportunity to build the most modern IP-based networks in the world.

- 'Disintermediation', which means removing the 'brokers' between two parties, is a major economic and social fallout of Internet technology. It also extends to the sphere of governance. Thus the Net can remove the layers of India's infamous opaque and corrupt bureaucracy and bring governance closer to citizens. Forms can be filled, taxes can be paid, notifications can be broadcast, citizens' views can be polled on controversial issues, the workings of different government committees can be reported and bills of government utilities can be recovered on the Net. Overall governance can become more citizen-friendly and transparent.

The scenario I am sketching is still futuristic even in the most advanced economies of the world. Further work is going on in three directions. One is in scaling up the present Internet backbone infrastructure to carry terabytes of data. The other is building IP-based routers to make it an efficient convergent carrier. And the third is to bridge the digital divide.

THE NEXT-GENERATION NET

"Bandwidth will play the same role in the new economy as oil played in the industrial economy till today," says Desh Deshpande, chairman, Sycamore Networks, a leader in Intelligent Optical Networking. He is part of the team that is working to create all-optical networks, optical switches and cross connects and 'soft optics'—a combination of software and optical hardware that can provision gigabytes of bandwidth to any customer in a very short period of time. "It used to take forever and lots of money to provision bandwidth for customers in the old telephone company infrastructure, but today technology exists to do it in a day or two. We want to bring it down to few minutes so that we can have bandwidth on demand," says Deshpande.

While there are several technologists like Desh Deshpande, Krish Bala, Rajeev Ramaswami, and Kumar Sivarajan who are working on improving data transport, Pradeep Sindhu is concerned with IP routers. Until the mid-nineties, very simple machines were being used as routers, and they received the packets, looked at them, sniffed them, and sent them off to the next router, but in the process wasted some time. They worked well at the enterprise level but when it came to several gigabytes of data in the core of the network, they could not handle it. Pradeep Sindhu was surprised at their primitive nature and proposed that computing had advanced enough by the mid-nineties to design faster and more efficient IP routers, and he built them for the core of the network. The company that Sindhu founded, Juniper Networks, has now come to be identified with high-end routers.

Sindhu has become an IP evangelist. "In 1996, when I asked myself how an exponential phenomenon like the Internet could be facilitated, I saw that the only protocol that could do it is IP. Since it is a connectionless protocol, it is reliable and easily scalable. The elements that were missing were IP routers. When I looked at the existing routers built by others, I was surprised at their primitive nature. That is when I realised that there was a great opportunity to build IP routers from the ground up using all the software and hardware techniques I had learnt at Xerox PARC (Palo Alto Research Center). I called Vinod Khosla, since I had done some work with Sun, and he had investments in networking. He gave me an hour. I spoke to him about the macro scene and told him that if we design routers from first principles, we could do fifty times better than what was available. He asked some questions and said he would think about it. He called back two weeks later and said let us do something together," reveals Sindhu.

"When Pradeep came to me, he had no business experience. My view was: 'I like the person and I like the way he thinks'. I asked him to sit next to somebody who was trying to build an Internet network for three weeks and asked him to understand what the problems are. He is such a good guy that he was able to learn quickly what the problems were. Helping a brilliant thinker like Pradeep and guiding him gives me great satisfaction. This is one guy who has really changed the Internet. The difference he has made is fabulous," says Vinod Khosla.

Khosla was one of the founders of Sun Microsystems in 1982 and has become a passionate backer of ideas to build the next generation Internet infrastructure. He works today as a partner in Kleiner Perkins Caulfield Byer, a highly respected venture capitalist firm in the Silicon Valley, and has been named by several international business magazines as one of the top VCs in the world for picking and backing a large number of good ideas.

BRINGING THE BYTES HOME

The second direction in which furious work is going on is to actually bring the bytes home. What is happening today is that a wide and fast Internet super highway is being built, but people still have to reach it through slow and bumpy bullock cart roads. This is called the problem of the 'edge' of the Net or the problem of 'last mile connectivity'.

While the 'core' of the network is being built with optical networking and fibre optics, several technologies are being tried out to reach homes and offices. One of them, laying fibre to the home itself, is expensive and can work in corporate offices in large cities. The second is to bring fibre as close to homes and offices as possible and then use multiple technologies to reach the destination. This is called fibre to the kerb. The options then available for last mile are:

- Using the existing copper wire cables of telephones and converting them to Digital Subscriber Lines (DSL). This will utilize the existing assets, but it works only for distances of about a kilometre, depending on the quality of copper connection.
- To use the coaxial cable infrastructure of cable TV. This requires a sophisticated cable network and not the one our neighbourhood *cablewallah* (cable service provider) has strung up.
- Use fixed Wireless in Local Loop. This is different from the limited-mobility services that are being offered, which are actually fully mobile technologies whose range is limited due to regulatory issues. Such mobile technologies are still not able to deliver bandwidths that can be called broadband.

However, fixed wireless technologies exist that can deliver megabytes of data. One of them is called Gigabit Wireless. According to Paul Raj at Stanford University, one of the pioneers in this technology, it can deliver several megabytes of bandwidth using closely spaced multiple antenna using a technique he developed called space time coding and modulation.

Another fixed wireless technology that is fast becoming popular as a way of building office networks without cables is Wireless LAN. This is being experimented with in neighbourhoods, airports, hotels, conference centres, exhibitions, etc as a way of delivering fast Internet service of up to a few megabytes and at least a hundred kilobytes. All one needs is what is called a Wi-Fi card in your laptop or desktop computer to hook onto the Internet in these environments.

The road outside my apartment has been dug up six times and has been in a permanent state of disrepair for the last two years. I have tried explaining that this is all for a bright future of broadband connectivity. Initially, they thought I was talking about a new type of cable TV and paid some attention to what I was saying, but their patience is wearing thin as roads continue to portray a Martian or lunar landscape and there is no sign of any kind of bandwidth, broad or otherwise.

But being an incorrigible optimist, I am ready to wait for new telecom networks to roll out and old ones to modernise, so that we will see a sea change in telecom and Internet connectivity in

India in a few years. In the process, I have learnt that if technology forecasting is hazardous then forecasting the completion of projects in India is even more so!

WHAT ABOUT VILLAGES?

Vinod Khosla, who does not mind espousing unpopular views if he is convinced they are right, says, "I suggest that we take our limited resources, and put them to the highest possible economic use. If you believe in the entrepreneurial model as I do, I believe that five per cent of the people empowered by the right tools can pull the remaining ninety-five per cent of the country along in a very radical way. The five per cent is not the richest or the most privileged or the people who can afford it the most; it is the people who can use it best. There are 500,000 villages in India. Trying to empower them with telecommunication is a bad idea. It's uneconomic. What we are getting is very few resources in the rural areas despite years of trying and good intent. There are sprawling cities, climbing to ten or twenty million people. And the villages lack power, communications, infrastructure, education, and health care. Talking about rural telephony to a village of 100 families is not reasonable. If we drew 5,000 circles, each 40 km in radius on the map of India, then we could cover 100 villages in each circle or about 100,000 in all. I can see a few thousand people effectively using all these technologies".

KNITTING THE VILLAGES INTO THE NET

However, Ashok Jhunjhunwala at IIT Madras disagrees. He believes that while it is a good idea to provide broadband connectivity to 5,000 towns, the surrounding villages can and should be provided with telephony and even intermediate-rate bandwidth. However, is there enough money for it and will there be an economic return? "Yes. Fixed wireless like the corDect developed at IIT Madras can do the job inexpensively", he asserts. "The point is to think out of the box and not follow blindly models developed elsewhere", he says. "Information is power and Internet is the biggest and cheapest source of information today. Thus, providing connectivity to rural India is a matter of deep empowerment", argues Jhunjhunwala.

What will be the cost of such a project? "Before we start discussing the costs, first let us agree on its necessity. Lack of access to the Internet is going to create strong divides within India. It is imperative that India acquire at least 200 million telephone and Internet connections at the earliest", adds he.

He points out that, today, telephony costs about Rs 30,000 per line. In such a situation, for an economically viable service to be rolled out by any investor, the subscriber should be willing to pay about Rs 1,000 per month. Jhunjhunwala estimates that only 2 per cent of Indian households will then be able to afford it. If we can, however, bring down the cost to Rs 10,000 per line, then 50 per cent of Indian households, which is approximately 200 million lines, become economically viable. "Breaking the Rs 10,000-per-line barrier will lead to a disruptive technology in India", says Jhunjhunwala.

LEARNING FROM THE CABLEWALLAH

Jhunjhunwala and his colleagues at Tenet (Telecom and Networking) Group at IIT Madras are working on this goal, but he feels that a much wider national effort is necessary to make this a reality. However, they have already developed a technology called corDect, which costs about Rs 8,000- Rs 12,000 per line and, more importantly, seamlessly provides, telephony and 30-70 Kbps Internet bandwidth. This is enough to start with, for rural networks. We can bring more advanced technology and greater bandwidth later. "We have a lot to learn from the neighbourhood *cablewallah*. From zero in 1992, the number of cable TV connections today is believed to have grown to over fifty million. What has enabled this? The low cost of a cable TV connection and the falling real value of TV. As a result, Cable TV has been made affordable to over sixty per cent of Indian households" says he.

"The second reason for this rapid growth", continues Jhunjhunwala, "is the nature of the organisation that delivers this service. Cable TV operators are small entrepreneurs. They put up a dish antenna and string cables on poles and trees to provide service in a radius of 1 km. The operator goes to each house to sell the service and collects the bill every month. He is available even on a Sunday evening to attend to customer complaints. This level of accountability has resulted in less-trained people providing better service, using a far more complex technology than that used by better-trained technicians handling relatively simple telephone wiring. However, what is even more important is that such a small-scale entrepreneur incurs labour cost several times lower than that in the organised sector. Such lower costs have been passed on to subscribers, making cable TV affordable".

BUILDING BRIDGES WITH BANDWIDTH

Tenet has worked closely with some new companies started by the alumni of IIT Madras, and the technology has been demonstrated in various parts of India and abroad. "It is possible to provide telephones as well as medium-rate Internet connections in all villages of India in about two years time with modest investment. We have orders of over two million lines of corDECT base stations and exchanges and about one million lines of subscriber units. Several companies like BSNL, MTNL, Reliance, Tata Teleservices, HFCL Info (Punjab) and Shyam Telelink (Rajasthan) have placed these orders. I think there is a potential of up to fifty million lines in India during the next four years. As for rural connectivity, n-Logue, a company promoted by the Tenet Group of IIT Madras, is already deploying Internet kiosks in villages in fifteen districts. The cost of a kiosk is about Rs 50,000. They are partnering with a local entrepreneur just as STD booths and cable TV did earlier and are providing rural telephony by tying up with Tata Indicom in Maharashtra and Tata Teleservices in Tamil Nadu. The local entrepreneurs can come up, villages can get telephony and the basic service operators can fulfill their rural telephony quota. It is a win-win solution", he says.

So there is reason for my optimism. More and more people in the government and private sector see the change that communication and Internet infrastructure can do to India, and the business opportunities available in it. In some places, the highway may be wider and at some others, narrower, depending on economic viability or present availability of capital, but one thing is sure: connectivity is coming to India in an unprecedented way. When this infrastructure is utilized innovatively, this nation of a billion people might see major changes in the quality of life taking place by 2020. Not only for the well off, but also for the hundreds of millions, if not the billion-plus others.

Bandwidth might bridge the real divides in India.

FURTHER READING

1. Where wizards stay up late-The origins of the Internet—Katie Hafner & Matthew Lyon, Touchstone, 1998
2. The Dream Machine-J C.R. Licklider and the revolution that made computing personal—M. Mitchell Waldrop, Viking 2001
3. Information Management: A Proposal—Tim Berners-Lee, CERN, March 1989, May 1990
4. Paul Baran, Interviewed by Judy O'Neill for the Oral History Archives, Charles Babbage Institute, Centre for the History of Information Processing, University of Minnesota, Minneapolis.
5. Making Telecom and Internet work for us—Ashok Jhunjhunwala, Tenet Group, IIT Madras
6. “The Semantic Web”—Tim Berners-Lee, James Hendler and Ora Lassila, *Scientific American,* May 2001
7. “WiLL is not enough”—Shivanand Kanavi, *Business India,* Sep 6-19, 1999
8. “Log in for network *nirvana*”—Shivanand Kanavi, *Business India,* Feb 7-20, 2000
9. “Pradeep Sindhu: A spark from Parc”—Shivanand Kanavi, *Business India,* Jan 22-Feb 4, 2001 (http://reflections-shivanand.blogspot.com/2007/12/profile-pradeep-sindhu.html)
10. “India: Telemedicine's Great New Frontier”—Guy Harris, *IEEE Spectrum,* April 2002

EPILOGUE

THE COLLECTIVE GENIUS

"The process of technological developments is like building a cathedral. "Over the course of several hundred years: people come along, lay down a block on top of the old foundations, saying, 'I built a cathedral.' Next month another block is placed atop the previous one. Then a historian asks, 'Who built the cathedral?' Peter added some stones here and Paul added a few more. You can con yourself into believing that you did the most important part, but the reality is that each contribution has to follow on to previous work. Everything is tied to everything else.
Too often history tends to be lazy and give credit to the planner and the funder of the cathedral. No single person can do it all, or ever does it all."

—**PAUL BARAN**, inventor of packet switching

Baran's wise words sum up the pitfalls in telling the story of technology. Individual genius plays a role but giving it a larger-than-life image robs it of historical perspective.

In India, there was a tradition of collective intellectual work. Take, for instance, the *Upanishads*,† or the *Rig Veda*;‡ no single person has claimed authorship of these works, much less the intellectual property rights. Most ancient literature is classified as *smriti* (memory, or, in this case, collective memory) or *shruti* (heard from others). Even Vyasa, the legendary author of the *Mahabharat*, claimed that he was only a raconteur. Indeed, it is a tradition in which an individual rarely claims "to have built the cathedral".

When I started researching this book, the success of Indian entrepreneurs in information technology was a well-known fact. As a journalist, I had met many of them, but I was curious to know which of them had contributed significantly to technological breakthroughs. While tracing the story of IT, I have also cited the work of several Indian technologists without laying any claim to completeness.

Nobody doubts the intellectual potential or economic potential of a billion Indians. However, to convert this potential into reality we need enabling mechanisms. The most important contribution of IT is the network. The network cuts across class, caste, creed, race, gender, nationality and all other sectarian barriers. The network, like all other collectives, creates new opportunities for collaboration, competition, commerce, cogitation and communication. It can inspire the collective Indian genius.

The hunger for opportunity, for knowledge, for change, is all there. I have seen it in the cities and villages of India. The political, intellectual and business elite of this country should break the barriers of the current networks of millions and build a network of a billion.

This is the call of the times: Hic rhodus, hic salta – *Here is the rose, now dance!*

---

†Ancient Hindu philosophical texts that summarise the philosophy of the Vedas. ‡Considered the oldest Hindu scripture, carried forward for centuries through oral tradition.